

Kant's Questions
What is the Human Being?

Patrick R. Frierson

Herr, was ist der Mensch, das du dich seiner annimmst?

– Psalm 144:3

The field of philosophy . . . can be reduced to the following questions: What can I know? What ought I to do? What may I hope? What is the human being? [*Was ist der Mensch?*] Metaphysics answers the first question, morals the second, religion the third, and anthropology the fourth. Fundamentally, however, we could reckon all of this as anthropology. (9: 25, cf. 11:249)

“I myself am a researcher by inclination. I feel the entire thirst for cognition and the eager restlessness to proceed further in it, as well as the satisfaction at every acquisition . . . [but] I would feel by far less useful than the common laborer if I did not believe that this consideration could impart a value to all others in order to establish the rights of humanity.”

— Immanuel Kant, from private notes written in 1764-5 (20:44)

Table of Contents

Acknowledgments

Introduction

Part One: Kant on the Human Being

- 1 Kant's Transcendental Anthropology
- 2 Kant's Empirical Anthropology
- 3 Human Evil and Human History
- 4 Human Diversity
- 5 Kant's Pragmatic Anthropology

Part Two: Responses and Alternatives to Kant

- 6 The Reception of Kant's Anthropology
- 7 19th Century Alternatives: Marx, Darwin, Nietzsche, & Freud

Part Three: What is the Human Being Today?

- 8 Scientific Naturalism
- 9 Historicism and Diversity
- 10 Existentialism
- 11 Normativity

Conclusion

Acknowledgements
XXX TO BE COMPLETED XXX

INTRODUCTION

[W]hat is man's ultimate nature? We keep returning to the subject with a sense of hesitancy and even dread. For if the brain is a machine of ten billion nerve cells and the mind can somehow be explained as the summed activity of a finite number of chemical and electrical reactions, boundaries limit the human prospect—we are biological and our souls cannot fly free. If humankind evolved by Darwinian natural selection, genetic chance and environmental necessity, not God, made the species . . . However much we embellish that stark conclusion with metaphor and imagery, it remains the philosophical legacy of the last century of scientific research. No way appears around this admittedly unappealing proposition. It is the essential first hypothesis for any serious consideration of the human condition. (Wilson 1978/2004: 1-2)

If we want to discover what the human being amounts to, we can only find it in what human beings are: and what human beings are, above all other things, is various. It is in understanding that variousness – its range, its nature, its basis, and its implications – that we shall come to construct a concept of human nature . . . To be human here is thus not to be Everyman; it is to be a particular kind of human being, and of course human beings differ . . . [I]t is in a systematic review and analysis of [different ways of being human] – of the Plain's Indian's bravura, the Hindu's obsessiveness, the Frenchman's rationalism, the Berber's anarchism, the American's optimism – that we shall find out what it is, or can be, to be a [hu]man. (Geertz 1973: 52-3)¹

[T]here is at least one being [the human being] whose existence comes before its essence, a being which exists before it can be defined by any conception of it. . . . What do we mean by saying that existence precedes essence? We mean that man first of all exists,

¹ Throughout this quotation and the following one, I've changed "man" and "men" here to "the human being" or "human beings."

encounters himself, surges up in the world – and defines himself afterwards. If the human being . . . is not definable, it is because to begin with he is nothing. He will not be anything until later, and then he will be what he makes of himself. Thus, there is no human nature The human being simply is. Not that he is simply what he conceives himself to be, but he is what he wills, and as he conceives himself after already existing – as he wills to be after that leap towards existence. The human being is nothing else but what he makes of himself. (Sartre 1993: 15)

What is the human being? The three quotations with which this introduction begins lay out three alternatives: E.O. Wilson, one of the preeminent sociobiologists of the 20th century, sees the answer to the question, or at least the parameters for the answer, given by biology. Humans are animals with a particular structure that has evolved over millions of years. We are biological beings, and what we need in order to better answer the question “What is the human being?” is better biology, a more detailed description of how we, as humans, are like and unlike other animals that inhabit the earth. Clifford Geertz, five years earlier, articulated a different conception of human beings, one at the core of the “human” science of anthropology. For Geertz, there is no answer to the question “What is *the* human being?” because we are not preeminently *biological* organisms, but *cultural* ones;² since there is not one human culture, there is not one kind of human being. What we need, to answer the question, is not better biological description, but more widely-ranging, deeply-investigating anthropological and historical studies of human variety. Rather than looking for a theory of human nature, we should seek a catalog of human ways of life. Alternatively, perhaps the problem of finding human nature is more difficult; perhaps, as the existentialist philosopher and literary author Jean-Paul Sartre argues, human beings are “condemned to be free” (Sartre 1956:568). Rather than trying to *discover* what human beings are, we should simply *make* human nature by free choices. Rather than looking as scientists or anthropologists at what human beings happen to be, we should take the role of architects of possibility, whether as artists (literary or otherwise) imagining and thereby creating new human possibilities, as political or social activists changing the social landscape, or simply as acting individuals creating human nature through our daily choices.

² To be fair to Wilson, this is a point with which he would, in most respects, agree.

The world we live in today is one within which these approaches to the question “what is the human being?” cannot be ignored. Scientific knowledge about our biological nature – from the coding of the human genome to the mapping of brain activity – has made it clearer than ever that humans operate with biological constraints. And as Wilson rightly points out, our knowledge of precisely how our biology is like and unlike that of other animals cannot be ignored in any serious consideration of human nature. At the same time, as the world becomes increasingly interconnected, human diversity, even if actually diminishing, is becoming more apparent and more relevant to more and more people across the globe. Protestant Christians in the United States cannot afford to be wholly ignorant of the cultures of rural Muslims in Afghanistan or atheists in China or Buddhists in Sri Lanka. Throughout the world, diversity is literally on one’s doorstep, as Catholic Filipinos work in Korea and Dubai, Muslim immigrants serve in European Parliaments, and Chinese businessmen set up shop in Africa. And the awareness of this diversity requires dealing with the fact that human nature is diverse. Finally, the increased power over our world and ourselves that comes from scientific, technological, and economic progress along with the awareness of the range of human possibilities that comes from seeing other cultures gives rise to an ever-more-acute sense that human nature really is up to us, that we can make ourselves into whatever we want to be. The situation in which human beings find themselves today requires thinking carefully about what it means to be human.

But what, precisely, is one seeking when one asks “What is the human being?” What is the question that Wilson, Geertz, Sartre, and so many others are trying to answer? Strikingly, *none* of these thinkers – not even the biologist Wilson -- treats the question “What is the human being?” quite like the question “what is oxygen?” or “what is a giraffe?” All of them see the question as one about our *prospects*, as one not merely about the structure of our brain or society, but about the *implications* of that structure for human choices, for what we should do with ourselves. And all recognize that the question “what is the human being?” is also, and fundamentally, about what is *important* about us.

When we understand it in this way, we can see why this question was central for Kant, why Kant would insist, “[t]he greatest concern of the human being is to know how to properly fulfill his station in creation and to rightly understand what one must do in order to be a human being” (20:41). Knowing what it is to be human is – for Wilson, Geertz, and Sartre no less than for Kant – something worthy of the greatest *concern*. Thus Geertz does not simply assert that humans are different, but adds that their differences are *more important* than their similarities, more

essential to what it means to be human. All of them recognize a point made by the early 20th century philosopher Martin Heidegger, that “The being whose analysis our task is, is always we ourselves. The being of this being is always *mine*” (Heidegger 1953:39). Or, put another way, they recognize what *anyone* who asks “what is the human being?” recognizes, that asking what a human being is really amounts to asking “who am I?”, “what is most important about me (to me)?”, “what do I value about myself?” and even “what do I aspire to be?”

This emphasis on values and aspirations, however, should not blind us to the fact that claims about human prospects and aspirations include descriptions of human beings. Even Sartre, who insists that what human beings *are* can only be answered *after* we make ourselves what we are to be, nonetheless recognizes that we are “condemned to be free”, that freedom is a “human condition” from which we cannot escape. Descriptions of the human condition provide the backdrop for his claims about how humans should act in response to them. Kant, too, recognizes the importance of accurate descriptions of human beings. In part, this is for practical reasons: “The question is which condition suits the human being, an inhabitant of the planet that orbits the sun at a distance of 200 diameters of the sun. Just as little as I can ascend from here to the planet Jupiter, so little do I demand to have qualities that are proper only to that planet . . . I do not at all have the ambition of wanting to be a seraph; my pride is only this, that I am a human being” (20:47). One needs to know what human beings *are* to know what we should aspire to be. And for Kant – as, at least, for Wilson and Geertz – human beings are also just very interesting to study. Even at a wholly natural and impersonal level, studying human diversity can be quite entertaining and humans’ biological and psychological nature is complex and challenging to explain. Not only must any practical account of human beings reflect an accurate description of them, but such descriptions are, in their own right, worth pursuing.

At its core, the question “What is the human being?” combines careful description of human characteristics with a normative, aspirational account of what about “us” is or would be truly valuable, an account rooted in the sense that each human questioner has of herself. Answering the question, however, involves clarifying what precisely constitutes a legitimate sort of “description” and also what structure and importance to ascribe to the normative, from-within perspective on oneself. And ultimately, as we see in the brief references to Wilson, Geertz, and Sartre, the answer to the question will combine – either implicitly or explicitly – these two aspects. The main purpose of this book is to lay out Kant’s answer to his question and to situate this answer in

the context of contemporary debates about human nature and the historical forces that brought us to where we are today. The first part of the book thus focuses on Kant's answer. The second part lays out the historical reception of Kant's ideas and the later trends that took the question in different directions. And the final part brings Kant into dialogue with the most important contemporary approaches to human nature, including those of Wilson, Geertz, and Sartre.

1. Kant's "Anthropologies"

In one of his lectures, Kant is recorded as having laid out his view of philosophy as a whole:

The field of philosophy . . . can be reduced to the following questions: What can I know? What ought I to do? What may I hope? What is the human being? Metaphysics answers the first question, morals the second, religion the third, and anthropology the fourth. Fundamentally, however, we could reckon all of this as anthropology. (9: 25, cf. 11:249)

The term "anthropology" may seem odd here for contemporary readers. We are accustomed to thinking of "anthropology" as a specific academic discipline that studies variations between people in different cultures. Kant, by contrast, uses the term anthropology in its original sense, as the study (*logos*) of human beings (*anthropos*). Thus Kantian anthropology includes comparisons between different people at different times, but it also includes – and emphasizes – general features of human beings as such. Anthropology is simply the discipline that answers the question, "What is the Human Being?" That is how the term will be used throughout this book.

But the claim that all of philosophy can be reckoned as anthropology may seem strange for other reasons, as well. While human nature may be a *part* of philosophy, philosophy often deals with other questions – such as the existence of God or the basic nature of reality – that seem to go beyond anthropology, and other disciplines deal (arguably even better) with important aspects of the answer to the question of the human being. In identifying philosophy and anthropology, Kant explicitly claims that every really important question that humans can ask, whether about God or substance or basic laws of physics or morals or aesthetics, is fundamentally a question *about* human beings, about what we can know, or should do, or may hope.

A final reason that Kant's claim to reduce all philosophy to anthropology might seem strange, especially for those accustomed to think of anthropology as an empirical discipline, is that this sort of enquiry seems inadequate to establish the normative claims embodied in the questions of what one *can* (legitimately) believe, or *should* do. Those familiar with Kant's work may be even more puzzled. At the end of his life, Kant did publish a book entitled *Anthropology from a Pragmatic Point of View*, but this book could hardly be said to include Kant's most important contributions to the questions of human knowledge, obligation, and hope. This *Anthropology* is striking for being deeply empirical, while Kant's most developed answers to the questions of knowledge, obligation, and hope emphasize that these questions must be answered non-empirically, or a priori. In his *Groundwork*, Kant even goes so far as to emphasize a distinction between "pure moral philosophy," which most fundamentally addresses the question "What ought I to do?" and "moral anthropology," which is secondary and merely adds empirical details. Kant's *Anthropology from a Pragmatic Point of View* relates to this secondary, empirical aspect of human knowledge, obligation, and hope, and does not articulate the most important dimensions of Kant's answers to the three questions that, supposedly, can all be "reckon[ed] . . . as anthropology."

In fact, Kant articulates different "anthropologies," different kinds of answer to the question "what is the human being?" Most importantly, he distinguishes between three ways in which one can ask the question and three dimensions of human life to which each of these three ways apply. The dimensions of human life arise from Kant's description of human mental states as being essentially of three kinds: cognitions (of truth), feelings (of pleasure), and volitions (for various goods). Kant does not ascribe consistent names to his three ways of inquiring, but in this book, I refer to them as "transcendental," "empirical," and "pragmatic." Put very briefly, transcendental anthropology provides normative, from-within accounts of what it's like to be human, accounts that define how one *should* think, feel, and choose based on what we take ourselves to be doing when we engage in thinking, feeling, or choosing. Empirical anthropology provides scientific (in a loose sense), observation-based descriptions and categorizations of how observable humans think, feel, and act. And pragmatic anthropology puts these two approaches together, drawing on empirical descriptions to provide advice about how best to satisfy the norms elucidated within transcendental anthropology. Part One of this book unpacks these different Kantian "anthropologies."

2. Receptions and Alternatives to Kant

Kant was not the first to ask about the human being, and he was certainly not the last. In the decades following the publication of Kant's three *Critiques*, Kant quickly rose to prominence and elicited significant responses to and criticisms of his views. But almost as quickly, Kant's specific approaches to thinking about human beings were overshadowed. At first, Kant was overshadowed by philosophers explicitly taking up his key themes, then increasingly by quite different philosophies like that of Hegel, and finally, over the course of the 19th century, by reflections on human nature by the biologists, psychologists, social scientists, literary critics and artists. This shift can be represented by the ways in which Darwin, Freud, Marx, and Nietzsche now dramatically overshadow Kant in discussions about human beings. Each of these figures not only developed important alternatives to Kant's conception of the human being but helped shift the discussion away from philosophy and into other arenas of inquiry. The relatively short middle section of the book traces these responses and alternatives to Kant's conception of the human being. Chapter six begins with the initial reception of Kant's transcendental anthropology and traces the series of criticisms and appropriations of Kant that culminated in Hegel's philosophy. Chapter seven then turns to Darwin, Freud, Marx, and Nietzsche not only for the ways in which they directly or indirectly challenge Kant but more importantly for the alternative visions of the human being that they present.

3. What is the human being today?

In the end, however, Kant's approach to the human being cannot satisfy "the greatest concern of the human being" (20:41) nor achieve the great goal that Kant assigned for it – "to establish the rights of humanity" (20:44) – unless it can be brought into conversation with the dominant approaches to thinking about human beings today. Part Three – chapters eight through eleven – cultivates this conversation. Each of these chapters describes one of the most important contemporary ways of answering the question "What is the human being?", and in each case, this answer is brought into dialogue with Kant.

Chapter eight looks at scientific naturalists such as Wilson, who advocate that the question is best answered by biological or psychological studies of human beings. There are a wide range of such naturalist approaches, so this chapter gives only a relatively small sample of the

ways in which philosophers and scientists have sought to use biological or psychological descriptions of human nature – what Kant would call “empirical anthropology” – to *fully* answer the question “What is the human being?” Chapter nine looks at approaches to human beings that emphasize human diversity, whether in the context of historical changes that make the human being of today different from the human beings of other times or in the context of contemporaneous cultural differences that make human beings in one culture different from those in another. Both of these approaches represent attempts to make what for Kant is only empirical anthropology (or even a subset of empirical anthropology) into the whole, and both approaches not only raise serious problems for Kant but also – as I hope to show – suffer from serious weaknesses that Kant’s anthropology can highlight and alleviate.

Among the most important weaknesses of historicism and naturalism is their failure to take sufficiently seriously what I will call the from-within perspective of transcendental anthropology, and chapter ten looks at a philosophical approach to human beings that seeks to take this quite seriously, but with a different result than Kant: existentialism. Existentialists are arguably the most direct *heirs* of Kant’s work in transcendental anthropology, taking a core insight of Kant’s – that human beings are fundamentally free but finite beings – and radicalizing this insight in such a way that the normative weight Kant ascribes to principles of reasoning and action becomes subordinated to – rather than constitutive of – human freedom.

In the end, I argue that while existentialism can reinvigorate and even enrich certain Kantian emphases, it fails to really speak to human beings because it fails to provide the right sort of normativity. Chapter eleven, then, takes up a small sample of contemporary approaches to normativity, beginning with some that are far removed from Kant and ending with two of the most prominent contemporary neo-Kantian philosophers writing today: Jürgen Habermas and Christine Korsgaard. These philosophers provide models for how Kantians today can integrate and respond to the insights of naturalism, historicism, and existentialism while still developing authentically Kantian conceptions of the human being.

PART ONE: KANT ON THE HUMAN BEING

CHAPTER 1: KANT'S TRANSCENDENTAL ANTHROPOLOGY

“As to the subject matter with which we are concerned, we ask that people think of it . . . as the foundation of human . . . dignity. Each individual equally, then, may reflect on it himself . . . [Our work] claims nothing . . . beyond what is mortal.”

— Francis Bacon, *New Organon*, quoted by Kant in the *Critique of Pure Reason* (Bii)

I. “Transcendental Anthropology”

In the introduction, I claimed that Kant’s answer to the question “What is the human being?” has at least three different components. Of these, I will refer to the one that made Kant famous and that he identified with “the field of philosophy” (9:25) as “transcendental anthropology.” The *term* “transcendental anthropology” is taken from Kant’s handwritten notes, in which he refers to an “anthropologia transcendentalis,” a “self-knowledge of the understanding and reason” that would critique all other sciences, including not only “geometry” and “knowledge of nature” but even “literature . . . theology, law” and “knowledge of morality” (RA 903, 15:395). But the concept of transcendental anthropology comes from combining Kant’s insistence that all of philosophy is reducible to “anthropology” (9:25) with his description of each aspect of his philosophy as “transcendental” (see A13/B27; 4:390; 5:113, 266, 270; 6:272; and 8:381).³ Throughout his philosophical works, Kant answers central philosophical questions in ways that are “anthropological,” but only in a new, distinctive sense of anthropology, one that I call “transcendental.”

³ Admittedly, Kant often reserves the term “anthropology” for his *pragmatic* anthropology, and he often reserves the term “transcendental” for investigations of the conditions of possibility of *experience* (the topic of his first *Critique*). His inclusion in “anthropology” of questions like “What may I know?” and “What ought I to do?”, questions most obviously discussed in Kant’s *Critiques* more than in his *Anthropology from a Pragmatic Point of View*, justifies referring to Kant’s “Critical” philosophy (named for his *Critiques*) as an “anthropology,” and all of his *Critiques* are “transcendental” in the sense laid out in the rest of this section (a priori, normative, from-within investigations of conditions of possibility). But the term “transcendental anthropology” is fundamentally a term used to contrast this approach to the human being with the empirical and pragmatic approaches used later in the book. It is not a term typically used by Kant to describe this approach.

While this transcendental investigation is contrasted, for Kant, with empirical study of human beings, one must be careful not to confuse “transcendental” with “transcendent” and thereby take transcendental anthropology (or philosophy) to refer to some aspect of human beings that transcends ordinary experience, or our animal nature, or something of that sort. In the same way that God might be seen as ultimately transcendent, we might want to study the transcendent aspect of human beings, through art, perhaps, or by talking about our immortal souls. Kant, however, sharply contrasts his transcendental philosophy from traditional philosophies of the “transcendent.” For Kant, “transcendental anthropology” is a kind of “self-knowledge of the understanding and of reason” (RA 903, 15:395). By this he does not mean simply that in knowing human beings, we know ourselves, since this would be true for empirical investigations of human beings as well. Instead, in transcendental anthropology, one knows oneself from-within rather than looking at one’s psychology from the stance of an observer. Transcendental anthropology is a most *immanent* sort of self-knowledge, and hence sharply contrasted with both empirical sciences *and* divine-like transcendence.

The notion of transcendental anthropology as “from-within” is often described in terms of a difference between “first person” and “third person” perspectives, the perspectives of the thinking, feeling, or choosing *subject* and perspectives *on* someone as an object. This way of describing the distinction can be helpful if one avoids thinking of “introspective” states as first person, since “from-within” does not imply that transcendental anthropology is “introspective” in any traditional sense. One way of making this distinction clear can be seen in the case of choosing a course of action. One observing humans might say that what a person chooses in a particular case is determined by accidental environmental features of which the person is only barely conscious. Or one might introspect and say that one’s behavior in a particular instance was caused by, say, a combination of anger and exhaustion. The next chapter shows how Kant’s *empirical* anthropology focuses on these sorts of causal explanations of behavior. But when one actually choosing, one doesn’t consider these accidental and unconscious influences as bases for choice. One looks for various reasons for action, and even if these reasons include what one might in another context see as mere causes of action (say, one’s desiring something), they have a different character when one considers them to be reasons to act; they serve not as *explanations* for behavior but as *justifications* for it. From-within the context of deliberation, one’s anger appears not as a necessary cause of action, but as a candidate reason for acting, a reason that one may either

endorse or reject. Throughout his transcendental anthropology, Kant offers accounts of what is involved from-within the processes of thinking, judging, choice, and aesthetic appreciation.

The from-within perspective involves an important evaluative or normative dimension. When explaining behavior non-transcendentally, one looks at what the causes of action are, and one need not evaluate whether these causes are “good.” The question whether, say, anger is a “good” *cause* of an action seems misguided; it either is the cause or it is not. But when thinking about behavior (or judgments, or choices) transcendently, one looks at reasons for that behavior, and reasons invite evaluation. Anger might have caused the behavior, but we can still ask whether it was a good reason for doing what one did. And this is the sort of question one asks, not merely when deciding what to do, but also when deciding what to believe, or how to judge about something, or even whether something is beautiful. The normative question – “Is this a good reason for people to do/think/feel such-and-such?” – arises within transcendental anthropology.

Along with this from-within, normative perspective on human beings, Kant’s transcendental anthropology employs a distinctive style of argument. “Transcendental” arguments in Kant proceed from some “given” to the conditions of possibility of that given. Thus Kant’s *Critique of Pure Reason* is an extended argument exploring the conditions of possibility of empirical cognition (what we can know). As an experiencer of the world, one can think about what must be the case for one’s experience to be possible, and Kant argues that in order for humans to have the kind of experience that we have, the world must contain substances, laws of causality, and other features, and human cognition of it must be limited in various ways. Similarly, the *Critique of Practical Reason* argues from the moral law we find valid within deliberation and evaluation to various conditions of possibility of that validity.

In sum, Kant’s *transcendental* anthropology focuses on what can be known about human beings *a priori* through an examination of their basic mental faculties “*from-within*” that specifically attends to the *conditions of possibility* of *normative* constraints on human beings. In the rest of this chapter, I take up some details of this transcendental anthropology as it plays out in Kant’s three famous *Critiques of Pure Reason*, *Practical Reason*, and *Judgment*. Before turning to those details, it is worth saying a bit more about the specifics of Kant’s conception of the human being in order to see how the *Critiques* hang together as a whole “transcendental anthropology.” Within both his empirical and his transcendental anthropology, Kant argues for a threefold division of

human mental states into those of cognition (thoughts), volition (desires and choices), and affection (feelings). Each different aspect of human beings is governed by its own a priori principles that are prescribed by a distinct higher cognitive power (5:196). In the *Critique of Judgment*, looking back on his philosophy as a whole, Kant uses a chart to show how his entire transcendental philosophy can be understood in terms of these different human faculties (5:198).⁴

Core aspect of the human being	Cognitive power that prescribes principles for it	A priori principles	Application to	Relevant Critique	Relevant Question
Cognition	Understanding	Lawfulness	Nature	<i>Critique of Pure Reason</i> (1781/1787)	What can I know?
Feeling	Judgment	Purposiveness	Art	<i>Critique of Judgment</i> (1790)	What may I hope?
Desire/Volition	Reason	Final End	Freedom	<i>Critique of Practical Reason</i> (1788)	What ought I to do?

II. What can I know?

The Critique of Pure Reason as transcendental anthropology of cognition

⁴ The first four columns are taken directly from Kant's own work, though I've edited them and modified terminology a bit. I've added the final two columns to show the connection with Kant's writings and his central questions. As with most of Kant's tidy charts, this one hides many complications (for instance, Kant typically identifies his question "what may I hope?" with his writings on religion and not directly with the *Critique of Judgment*), but it is helpful for a general overview.

Kant's most famous and important work, the *Critique of Pure Reason*, was published in 1781 after a "Silent Decade" during which he published virtually nothing. The work was and is his magnum opus, the work that defined him as a philosopher. Its specific focus is anthropological in that it focuses on a particular human capacity: "getting to the bottom of the faculty we call the understanding and . . . the determination of the rules and boundaries of its use" (A xvi). However, Kant is not interested here in the empirical question of how the understanding operates, but in giving an account of the rules under which it must operate and the limits that these rules imply for how far we should seek to extend our knowledge. In the process, Kant aims to answer the question "What can I know?" as that question applies specifically to the "objective validity" of "a priori concepts" (A xvi), that is, "what and how much can the understanding and reason cognize free of all experience?" (A xvii). Through this transcendental anthropology of cognition, Kant defends a metaphysics that consists in a priori claims about the nature of the world and lays out an epistemology that limits the scope of such claims.

For Kant, metaphysics involves "a priori synthetic" claims. An *a priori* claim is one that is universal and necessary and thus not based merely on empirical generalizations. Many a priori claims, however, such as the claims that "bachelors are unmarried" or "a=a" merely unpack concepts and do not tell anything substantive about the world. Kant calls these sorts of empty claims "analytic" because they merely "analyze" concepts, showing what is already contained within them. Other claims, by contrast, are "synthetic" because they synthesize, or put together, concepts that, in themselves, are distinct. The claims that "Hondas are reliable" or "honey is sticky" are synthetic; reliability and stickiness are not part of the *definitions* of Hondas and honey. Of course, those claims are also empirical, rather than a priori. But there are some claims, such as "every change has some cause," that are both synthetic and a priori. Here one does not merely define changes as having causes (because then one could ask whether the ripening of fruit is really a "change" if there is no obvious cause) but makes a contentful claim about all changes in the world. But this claim is not based on empirically generalizing our experience that all changes have causes. We have not actually observed causes for every change; and were someone to claim that a particular change lacked a cause, we would insist that the cause had simply not yet been discovered.⁵ When we claim that every change has some cause, we

⁵ The development of modern quantum mechanics both problematizes and confirms this claim. On the one hand, the dominant Copenhagen interpretation of quantum mechanics *does* posit uncaused changes, or at least changes that are undetermined by their causes. On the other hand, Einstein's reaction to this

claim a necessary connection (hence, a priori) between distinct concepts (hence, synthetic). The need for a metaphysics that is at once synthetic and a priori raises “the general problem” of the *Critique of Pure Reason*: “How are synthetic judgments possible a priori?” (B19, cf. *Prolegomena to any Future Metaphysics*, 4:276).

Kant’s answer to this question depends upon conceiving of metaphysics as a subset of transcendental anthropology. From the beginning of his *Critique*, Kant makes clear how radically human-centered his metaphysics is, comparing the fundamental shift in thinking embodied in this *Critique* to the revolution in astronomy effected by Copernicus:

Up to now it has been assumed that all our cognition must conform to objects; but all attempts to find out something about them a priori through concepts that would extend our cognition have, on this presupposition, come to nothing. Hence let us once try whether we do not get farther with the problems of metaphysics by assuming that the objects must conform to our cognition, which would agree better with the requested possibility of an a priori cognition of them, which is to establish something about objects before they are given to us. This would be just like the first thoughts of Copernicus, who, when he did not make good progress in the explanation of the celestial motions if he assumed that the entire celestial host revolves around the observer, tried to see if he might not have greater success if he made the observer revolve and left the stars at rest. (Bxvi)

Kant’s reversal is as radical as Copernicus’s but moves in the opposite direction. Astronomers before Copernicus thought of the earth – and thus human beings – as the center of the universe. Copernicus’s radical shift in perspective, describing the earth as just one of many planets circling the sun, moves humans out of the center of the universe.

By contrast, Kant moves human cognition *into* the center of metaphysics. He begins by isolating an assumption of prior metaphysics, the assumption that in order to know anything about the world, our judgments about the world have to conform to the way the world really is. Kant claims that this assumption has made progress in metaphysics impossible. Broadly speaking, previous philosophers – especially during

interpretation was precisely to argue that its proponents had simply failed to look hard enough to find the relevant “hidden variables” (see Einstein 1935 and Bohr 1935). I discuss the relationship between contemporary physics and Kant’s transcendental anthropology of cognition in chapter nine.

the 17th and 18th centuries⁶ – were either “rationalists” who sought philosophical systems based upon reason alone or “empiricists” who sought the ultimate foundations of knowledge in experience. But empiricists fail to account for the aprioricity of metaphysics, while rationalists fail to properly account for its synthetic status (by mistakenly overestimating what reason alone can do). Kant’s Copernican turn is based on the thought that empiricists and rationalists fail because both are looking for a way to make human cognitions fit onto an independently given world of objects. There is better hope of showing how a priori synthetic judgments are possible if one assumes instead that the world of objects must conform to the structure of human cognition.

Let us once try whether we do not get farther with the problems of metaphysics by assuming that the objects must conform to our cognition, which would agree better with the requested possibility of a priori cognition of them, which is to establish something about objects before they are given to us. (B xvi)

By assuming that the world must conform to our cognition, it is possible to have knowledge of the world based on the structure of our cognition rather than by induction from what we experience. By itself, this claim does not get beyond merely analytic claims. The world must conform to my cognition in that bachelors in the world must be unmarried, but to get a priori substantive claims, Kant must do more.

Kant’s next move both limits the scope of this Copernican turn and helps show how it functions to make substantive (or “synthetic”) a priori knowledge possible. Kant claims that human cognition has both a passive and an active component.

Our cognition arises from two fundamental sources in the mind, the first of which is the reception of representations (the receptivity of impressions), the second the faculty for cognizing an object by means of these representations (spontaneity of concepts); through the former an object is given to us, through the latter it is thought in relation to that representation . . . If we call the receptivity of our mind to receive representations insofar as it is affected in some way sensibility, then . . . the faculty for bringing forth representations itself, or the spontaneity of cognition, is the understanding. It comes along with our nature that intuition [that through which cognition relates immediately to an object] can never be other than sensible . . . The

⁶ In his “History of Pure Reason,” Kant traces the distinction between empiricists and rationalists back to Aristotle and Plato (see A854/B882), but it is most commonly associated with the major empiricists (Locke, Berkeley, and Hume) and rationalists (Descartes, Spinoza, and Leibniz) of the early modern period.

faculty for thinking of objects of sensible intuition, on the contrary, is the understanding . . . Thoughts without content are empty, intuitions without concepts are blind. (A50-1/B74-5, cf. A19-20/B33-4)

Despite the deluge of technical terminology, Kant's point here is simply that humans' thoughts about objects have two components, an active component by which we *think* about objects, and a passive component by which thoughts are *about objects*. When one merely thinks about relations between concepts, one makes use of an active capacity for thought, but this capacity is not directed towards any real objects. It is, as Kant says, "empty." And when one merely "takes in" the world without actually conceptualizing it, one does not really see what one is exposed to; one's taking-in of the world is effectively "blind." Knowledge of a real world involves receiving "impressions" from the world and processing them using one's concepts.

Kant's appeal to "intuitions"⁷ given through sensibility limits the scope of the Copernican turn. Kant does not claim, and need not claim, that *everything* about the empirical world is determined by the structure of human cognition. Because we have a receptive faculty, humans have knowledge we take *from* the world, such as the knowledge that there are mountains in the Pacific Northwest of North America, that water freezes, that dogs and cats cannot interbreed, that large material objects are made of small molecules, etc. And there are other claims that are false, but if true, would have to be discovered empirically, such as the existence of the Loch Ness monster, or fairies, or solid crystalline spheres rotating in the heavens. For such empirical knowledge, cognition must conform to the world. The world will not have fairies in it just because we believe in fairies, nor will it cease to have molecules if we cease to (or do not yet) believe in molecules. Kant's Copernican turn justifies the possibility of some substantive a priori knowledge of the world, but it does not justify claiming to know everything about the empirical world simply by reflecting on one's cognitive capacities.

But Kant also argues that the distinction between intuitions and concepts (and relatedly between sensibility and the understanding) provides for the possibility of a priori knowledge that goes beyond mere conceptual analysis because even our receptivity to the world has an a priori structure to which the world must conform.

⁷ Kant's notion of an "intuition" is quite different from our contemporary use of the term to refer to judgments of which one is immediately certain. For Kant, an intuition is "that through which [a cognition] relates immediately to objects" (A19/B33). An intuition is thus something like an immediate, pre-conceptual, sensory awareness of an object.

I call that in the appearance which corresponds to sensation its matter, but that which allows the manifold of appearance to be intuited . . . I call the form of appearance. Since that within which the sensations can alone be ordered and placed in a certain form cannot itself be in turn sensation, the matter of all appearance is only given us a posteriori, but its form must all lie ready for it in the mind a priori. (A20/B34)

Humans' capacities to be affected by the world have a particular structure, and it should be possible to develop an a priori science of the principles of this sensibility. Moreover, precisely because sensibility is a faculty of intuitions rather than of concepts, an a priori science of sensibility will not proceed simply by unpacking concepts, and thus may provide a way of justifying claims that are both a priori and synthetic.

Kant's approach is made clearer when he turns to the details, where he argues that space and time are the two a priori intuitions that structure all empirical intuitions. Both are a priori because we cannot think of the world as non-spatial or non-temporal and because we could never think of external objects without an already-given spatial structure nor of succession without an already-given temporal structure. Both are a priori intuitions, rather than concepts, because they are represented as given structures within which particular objects appear, rather than as constructed concepts under which objects fall.⁸ Kant reiterates the status of space and time as a priori intuitions by pointing out that these intuitions underlie the success of geometry (in the case of space) and arithmetic (in the case of time), both of which give synthetic a priori knowledge.⁹ In geometry, for example, one does not measure shapes to discern their properties empirically, but one also does not merely analyze the concepts of those shapes. Instead, one uses mental, spatial images (or axioms about mental space) to connect concepts not already united by definition. Thus, for example, we can know (without empirically measuring every triangle in the world) that the sum of the interior angles of any triangle makes two right angles, even though this fact is not contained in the definition of a triangle. As an intuition, space is able to ground synthetic claims. As an a priori intuition, it grounds these claims a priori. "Thus," Kant says, "our explanation alone makes the possibility of geometry as a synthetic a priori cognition comprehensible" (A25/B41).

⁸ For these arguments, see A22-5/B37-40 (space) and A30-2/B46-8 (time). The precise details of these arguments are controversial and the success of the arguments is contested. For discussion, see Allison 1983, Guyer 1987.

⁹ A25/B40-1.

The understanding, like sensibility, has an a priori structure, and the heart of Kant's *Critique of Pure Reason* focuses on showing that the a priori structure of the understanding is a necessary condition for the possibility of experience. Kant's argument for this claim is the most difficult part of the *Critique of Pure Reason*, and Kant revised the argument extensively between the first and second editions of the book.¹⁰ The essence of Kant's argument is twofold. First, he looks to "logical function of the understanding in judgments" for "the clue to the discovery of all pure concepts of the understanding" (A70/B95). The basic idea here is that once we are committed to seeing experience as a product of the understanding, one can examine how the understanding is used in constructing and relating *judgments* to discern fundamental categories of the construction and relation of *objects* within experience. Since the operation of the understanding with respect to judgments is the provenance of a relatively straightforward logic, one can simply use what Kant calls "general logic" as a clue to the transcendental anthropology of the understanding. For example, Kant argues that because all relationships within judgments must be either categorical (x is y), hypothetical (if x, then y), or disjunctive (x or y), the understanding must have corresponding a priori categories for cognizing objects. Respectively, these are: inherence and subsistence (cognizing objects as things with properties), cause and effect, and community (objects as reciprocally interacting). Overall, Kant lays out a famous (or infamous) table of twelve basic categories of the understanding, drawn from a similar table of twelve functions of thinking.

Of Quantity

Unity

Plurality

Totality

Of Quality

Reality

Negation

¹⁰ For very good detailed studies of Kant's key arguments here, see Allison 1983, Ameriks 2003, Guyer 1987, and Longuenesse 1997.

Limitation

Of Relation

Of Inherence and Subsistence (*substantia et accidens*)

Of Causality and Dependence (cause and effect)

Of Community (reciprocity between agent and patient)

Of Modality

Possibility – Impossibility

Existence – Non-existence

Necessity – Contingency

This table proposes only what the content of a priori categories of the understanding would be; Kant still needs to show that experience requires any such categories. In particular, given that objects are presented in terms of spatial and temporal structures of sensibility, one might wonder whether any further contribution from the understanding is needed. For that purpose, Kant lays out a detailed argument for the necessary role of the understanding in making cognition of objects possible. The key move in this argument is Kant's claim that experience is possible only by virtue of a twofold "unity," on the side of both the "object" and the "apperception" of that object. Kant takes the term "apperception" from Leibniz, for whom it referred to something like explicit consciousness of an object as opposed to mere unconscious awareness of it, the difference between hearing sounds when one is asleep and listening to them when one is awake. For Kant, "apperception" refers to "the 'I think' [that] must be able to accompany all of my representations, for otherwise something would be represented in me that could not be thought at all" (B131-2). Kant's basic idea in this section of the *Critique* is to connect experience of unified objects with a unified "I think" and thereby with the categories, as the principles of any such unity. First, Kant argues that in order for one to unite different representations together into consciousness of a single object, those different representations must be held together in a single

consciousness. To have a cognition of a purple cow, it will not do for one person to have a representation of the color purple and another to have a representation of a cow. These different representations, to be united into a cognition of a single object, must be united by a single consciousness.

[C]ognitions . . . consist in the determinate relation of a given representation to an object. But an object is that in the concept of which the manifold of a given intuition is *united*. Now, however, all unification of representations requires unity of consciousness. Consequently the unity of consciousness is that which alone constitutes the relation of representations to an object. (B 137)

The relevant unity of consciousness here is not that “subjective unity of consciousness which is a determination of inner sense” (B139), that is, it is not an introspective awareness of oneself as united throughout time. Rather, the relevant unity is the from-within unity by virtue of which one becomes conscious of and makes justifiable claims about objects in the world.

Now just as the table of judgment provided Kant’s “clue” to discovering a set of a priori categories, so he argues that we see, in the form of judgment itself, a necessary appeal to this transcendental unity: “a judgment is nothing other than the way to bring given cognitions to the objective unity of apperception. That is the aim of the copula *is* in them . . . For this word designates the relation of the representations to the original apperception and its [apperception’s] *necessary unity*, even if the judgment itself is empirical” (B 141-2). In other words, in making a judgment such as “mangos are delicious,” one not only makes an empirical claim about the world but implicitly asserts the necessary unity of the I that holds together “mangos” and “delicious” in a single “I think.” And now Kant comes to his punchline:

Therefore the manifold [of different representations], insofar as it is given in *one* empirical intuition, is *determined* in regard to one of the logical functions for judgment, by means of which . . . it is brought to a consciousness in general. But . . . the categories are nothing other than these very functions for judging, insofar as the manifold of a given intuition is determined with regard to them. Thus the manifold in a given intuition also necessarily stands under categories. (B 143)

Thus if, e.g., I make the empirical intuition of a house into perception through apperception of its manifold, my ground is the *necessary unity* of space . . . This same synthetic unity, however, if I abstract

from the form of space, has its seat in the understanding, . . . in the category of *quantity*. (B162)

Our “apperception” of objects is unified through organizing various representations together by means of a priori concepts of the understanding such as unity, reality, or causation. Even spatial and temporal properties of an object only become *properties* of an *object* when subsumed together by the understanding, under concepts such as “property (of a substance).” These a priori concepts, or categories, provide a from-within structure by means of which a human knower can connect representations in such a way that those representations remain part of a single “I think,” and they provide the framework according to which a mere “manifold” can become a coherent *object* of experience.

At this point, one clarificatory warning is needed lest one think that Kant’s transcendental anthropology does (or claims) more than it really does. Rene Descartes, too, emphasizes the importance of the “I think,” claiming that he could be absolutely certain of the claim, “I think therefore I am” and reasoning, even further, that the nature of the “I think” implies the simplicity, unity, and ultimately immortality of the human soul. For Kant, however, such inferences mistakenly treat the transcendental unity of apperception that makes (empirical) cognition possible as itself a possible *object* of (such) cognition. In his transcendental deduction, Kant emphasizes that while the “transcendental synthesis of the manifold” makes me conscious “*that I am*,” it reveals neither what “I am in myself” nor how “I appear to myself” (B157). Kant criticizes attempts to get from the formal requirements of the “I think” to substantive claims about such an I, calling such attempts “Paralogisms,” or invalid arguments (see A341/B399-A404/B431). For example, the inference from the necessity of a unified “subject” of thought to a unified thinking “substance” erroneously takes the formal category of “subject” to have objective meaning, when “pure categories (and among them also the category of substance) have in themselves no objective significance . . . unless an intuition is subsumed under them” (A348-9). Because one can have no intuition of the “I think” that unifies all intuitions, the only possible *objective* cognition of oneself is of “our own subject only as appearance,” and this is available “through inner sense” (B156). And Kant argues in a “Refutation of [Cartesian] Idealism” (B274-9) that knowledge of this empirical self is secondary to knowledge of external objections, since “the determination of my existence in time is possible only by means of . . . actual [persisting] things that I perceive outside myself” (B 275). Kant’s appeal to the transcendental unity of apperception is an explanation of cognition from-within, one that shows that objective empirical cognition requires unifying one’s representations

by means of a priori categories. It thereby shows that the objective world must be unified in that way, but it makes no objective claims about the apperception that unified that world.

Having laid out the a priori structure of both sensibility and the understanding, Kant turns to the way in which these two different cognitive faculties work together to structure the world of experience. By showing how humans' a priori categories work with sensibility to structure the empirical world, Kant's "system of all principles of pure understanding" provides the a priori metaphysics promised in his Preface. The specific details of the various ways in which these faculties combine is both complicated and contested, but one example (Kant's best known) is sufficient to give a sense for his general strategy. One of the principles that Kant defends as a principle by which human beings structure the objective world is that of cause and effect: "All alterations occur in accordance with the law of the connection of cause and effect" (B232). Kant's argument proceeds by considering what is necessary in order for a set of perceptions to be considered perceptions of alteration (or, more generally, of something happening). Kant distinguishes between merely subjective perceptions and objective experience. To have objective experience, one must organize perceptions in accordance with categories. But to have experience of objective alteration (succession), perceptions must be ordered in accordance with the category of cause—effect. If ordered using inherence—subsistence (seeing some perceptions as properties of others) or community (seeing each perception as part of a whole), the sequence of one's perceptions would not refer to an objective sequence, since objectively, one supposes that the properties of the thing exist at the same time as the thing and one supposes all the parts of a thing to exist at the same time.¹¹ Given the Copernican turn, to say that one must order perceptions in a certain way is just to say that the

¹¹ Kant gives, as an example of a purely subjective sequence, the perception of a house, starting with the roof, then the windows, then the door, and then the chimney. Here one doesn't suppose that objectively speaking there really is first a roof, then windows, then a door, etc. On the other hand, one might actually suppose that the order of one's perceptions does correspond to an objective order. For this, Kant gives the example of a boat. One perceives a boat upstream, a boat midstream, and a boat downstream, and one supposes not that these are different parts of a complicated stream-wide boat, but that in reality – that is, objectively – the boat is moving. Kant then considers what sort of concepts one would have to impose on one's set of perceptions to order them in such a way that one considers their order objective. His answer is that the perceptions would have to be thought of as though they have to occur in the order in which they do. And this necessary sequence of perceptions must be according to some rule. But necessary sequence according to a rule is just what one thinks of when one thinks of the relation between cause and effect. So if one is to think of the order of perceptions order as referring to an objective order, one must impose the concepts of cause and effect on those perceptions.

objects of those perceptions must in fact be ordered in that way. By imposing an aspect of the structure of the human understanding – the category of cause and effect – on the subjective flow of perceptions in inner sense, human beings are able to perceive the world (and thereby structure the world) as a series of causally determined changes.

Throughout his proofs – for the necessity of space, time, the categories, and causation – Kant does not provide merely empirical claims about human cognition. Consistent with his insistence on transcendental anthropology, Kant looks at cognition from-within, arguing that certain cognitive presuppositions are necessary conditions of the possibility of justifying the claims that we make about the world. Because, from-within, we take mathematics to be justified, we must assume that space and time structure our world. Because we can make justified empirical claims about objects, we must be organizing and unifying the diffuse manifold of intuition into coherent cognitions. And because some of this cognition is of objective succession, we must apply categories of cause and effect to structure the world we experience. In the end, Kant's *Critique of Pure Reason* provides a transcendental analysis of human faculties of sensibility and understanding that elucidates their a priori structure and the contributions of this structure to experience of an objective world. Human beings for whom experience and a priori synthetic judgments are possible are finite beings dependent upon sensibility and also spontaneous free thinkers. Moreover, given Kant's Copernican turn, this transcendental anthropology provides both an epistemology that delimits what we can know and a metaphysics, in that the world itself must conform to the structures of human cognition. Metaphysics and epistemology turn out, in Kant's hands, to be reckoned as (transcendental) anthropology.¹²

¹² The radicality of Kant's position here can be seen in one of the most famous criticisms of it. Bertrand Russell critiques Kant's account of a priori knowledge on the grounds that, among other things, Kant relies on contingent facts about human nature. As Russell explains,

The thing to be accounted for is our certainty that the facts must always conform to logic and arithmetic. To say that logic and arithmetic are contributed by us does not account for this. Our nature is as much a fact of the existing world as anything, and there can be no certainty that it will remain constant. It might happen, if Kant is right, that to-morrow our nature would so change as to make two and two become five. This possibility seems never to have occurred to him. (Russell 1912/1998: 87)

Russell's problem arises because Kant seems to ground the a priori necessity of truths of mathematics (and even logic) on the conditions of possibility of our sensing and thinking about the world. Because we humans perceive the world in Euclidian space and time and think about it using various logical categories,

With his analysis of the way in which sensibility and the understanding combine to structure a knowable empirical world and his defense of several specific a priori principles of human cognition to which that empirical world must conform, Kant completes the first part of his answer to the question “What can I know?” But Kant’s transcendental anthropology of cognition involves two further elements as well. One of these is not continued until a subsequent work. In his *Metaphysical Foundations of Natural Science*, Kant argues that “natural science presupposes . . . metaphysics of nature,” which includes not only the “laws that make possible the concept of a nature in general,” laid out in his *Critique of Pure Reason*, but also laws that “concern [themselves] with a particular nature of this or that kind of things, for which an empirical concept is given, but still in such a manner that, outside of what lies in this concept, no other empirical principle is used for [their] cognition” (4:469-70). The general idea is that the nature of the human mind is such that if it cognizes, say, material bodies, then it will have to cognize them in particular ways. Given the Copernican turn, these necessary ways of cognizing would be a priori synthetic principles of material bodies themselves. What this implies, for Kant, is that the basic principles of physics itself can be seen as a sort of transcendental anthropology of cognition. In *Metaphysical Foundations of Natural Science*, Kant argues that given the barest concept of matter in motion, one can derive a priori such claims as the conservation of matter (4:541), the claim that all motion is relative to a particular frame of reference (4:487), and Newton’s laws (e.g. that “every change in matter has an external cause” (4:543) and “in all communication of motion, action and reaction are always equal” (4:554)). For Kant, not only the most basic metaphysical claims about the universe, but even its most basic physical laws are a priori conditions of the possibility of experiencing nature. And they are a priori conditions precisely because they reflect transcendental structures of human cognition. For Kant, Newtonian physics is transcendental anthropology.

The second further element comes in the second part of the *Critique of Pure Reason*. There Kant turns from human sensibility and

truths of geometry, arithmetic, and logic must be true in any world that is a world for us. But, Russell suggests, that makes these truths ultimately contingent upon human nature, and a truth that is contingent upon human nature cannot be the sort of a priori – and hence necessary – truth that Kant sought in his *Critique of Pure Reason*. (See Frierson 2009; Guyer 2006: 66-67, 81-2; and vanCleve 1999: 36-40 for Kantian responses to Russell’s objection.)

understanding to human “reason.” The a priori forms of sensibility and the understanding are “constitutive” of experience; the objective world is constituted by conformity with these forms. Reason, by contrast, presents ideals that regulate humans’ pursuit of knowledge but that are not constitutive of that knowledge. In the context of this *Critique* reason can be understood as that most philosophical of cognitive powers, the one that constantly asks the question, “Why?” And in asking this question, reason constantly seeks the “unconditioned,” that is, an answer that does not itself require a further explanation. For Kant, this search for the unconditioned plays itself out in every area of knowledge. Reason prompts humans to seek the causes of phenomena in our world, then the causes of these causes, and so on. It prompts us to look for the constituent parts of the objects in our world, the constituent parts of those, and so on. It prompts us to develop an idea of an ens realissimum, a most real being whose perfection is not conditioned by anything. In these ways and more, reason drives human knowers to discover more and more about the world in which we live.

But with this regulative function of reason comes a dangerous illusion, one that “does not cease even though it is uncovered and its nullity is clearly seen into by transcendental criticism” (A297/B353, see too Avii). Reason drives humans to learn more and more about their world in a search for the unconditioned, but this impulse naturally generates the illusion that an unconditioned is there to be found. As we search for the smallest and most basic particle, or the first cause of the universe, or the most perfect possible being, reason incites us to assume that there is in fact a most basic particle, a first cause, or a perfect being. The second half of Kant’s *Critique of Pure Reason* focuses on showing why these assumptions are unwarranted. The basic argument in each case is that the conditions of possibility of objective cognition conflict with the nature of the unconditioned, such that these ideal goals of reason are the sorts of things that could never exist in a world structured by human forms of intuition and understanding. While we should still make use of reasons’ ideals as ideals guiding inquiry into the world of experience, we cannot treat them as really existing in that world.

The details of these arguments are unnecessary in this brief account of Kant’s transcendental anthropology, but it is worth looking at one of Kant’s specific discussions: the third “antinomy of pure reason.”¹³

¹³ Kant does not introduce this discussion in the context of a discussion of human freedom. Instead, the issue of freedom comes up in the context of thinking about whether, in general, “causality in accordance with the laws of nature is . . . the only one from which all appearances . . . can be derived” (A444.B472), but Kant quickly shifts to talking about the possibility of human freedom as a causality that grounds

In his antinomies, Kant shows that when one assumes that the ideals of reason must apply to the objects of possible experience, it is possible to prove contradictory positions on many of the most important questions of metaphysics. The third antinomy starts by proving both that “Causality in accordance with the laws of nature is not the only one It is also necessary to assume another causality through freedom” (A444/B472) and that “There is no freedom, but everything in the world happens solely in accordance with laws of nature” (A445/B473). The proof of freedom appeals to reason’s need to seek the unconditioned and argues that no explanation can be sufficient if conditioned by further explanation. The argument against freedom draws on Kant’s defense of universal causation in the second analogy (discussed above) to show that there could never be a cause that was not itself the effect of a further cause, since any act of causation would have to be an event in the world, and every event must have a cause.

Kant’s “solution” to the antinomy involves two components, both of which are important for understanding Kant’s transcendental anthropology. Kant first draws attention to the merely regulative function of ideals of reason. Insofar as the demand for sufficiency is a regulative ideal, the empirical investigation of natural causes should constantly seek a “sufficient” cause for any appearance by investigating causes as far back as they will go. But any cause that one in fact finds will, by virtue of being itself an appearance, call for investigation into its own cause. Thus we must *seek* sufficiency in our explanations, but we will never find it.

But Kant’s resolution of the third antinomy introduces a second aspect of his transcendental philosophy which is crucial to understanding Kant’s answer to the question “What is the human being?” So far, my discussion of Kant’s transcendental anthropology of cognition has focused on the *positive* contribution that this anthropology can make towards a robust metaphysics of nature: given that our empirical knowledge depends upon the structure of our sensibility and understanding, we can prove important claims about the empirical world a priori. But Kant points out that this positive contribution entails “a very strange result . . . , namely, that with this faculty [of cognition] we can never go beyond the boundaries of possible experience” (Bxx). We can establish a priori claims *about possible objects of experience*, but we cannot provide any theoretical justification for any claims about unexperienceable things. Kant calls such things “things-in-themselves” or “noumena” and distinguishes them from objects of possible

appearances in the world.

experience, which he calls phenomena. And throughout his transcendental account of cognition, Kant reminds his readers that the nature of human cognition determines only the way in which “objects” (of possible experience) must be, not the way in which “things-in-themselves” must be. Thus, for example, when he proves a priori that the world must be spatial-temporal, Kant claims,

Our expositions accordingly teach the reality (i.e. objective validity) of space [and time] in regard to everything that can come before us externally as an object, but at the same time the ideality of space in regard to things when they are considered in themselves through reason, i.e., without taking account of the constitution of our sensibility. (A28/B44)

For something to be an empirical object, it must be presented to us through sensibility. And since the structure of our sensibility is spatial, all empirical objects must be spatial. But precisely because the spatiality of empirical objects is due to our sensibility, we are not justified in saying that such objects are spatial apart from human sensibility. The result is that Kant’s metaphysics commits him to what he calls an “empirical realism” but a “transcendental idealism.” His metaphysics is empirically real because its claims (e.g. about causality) are necessarily true of the empirical world. But it is transcendently ideal because such claims are limited to the empirical world and say nothing about what “things-in-themselves” – apart from human sensibility – might be like.

When Kant turns to the third antinomy, this transcendental idealism does significant work. Consistent with the insights of his second analogy, Kant insists that any objective alteration must be the result of causes in accordance with natural laws. But he turns in his resolution to the third antinomy to the question of “whether it is a correct disjunction that every effect in the world must arise either from nature or freedom, or whether instead both, each in a different relation, might be able to take place simultaneously” (A336/B564). Given Kant’s transcendental idealism, the law-governed causality of the empirical world does not preclude a different kind of causality – freedom – operating at the level of things-in-themselves:

[F]or a subject of the world of sense we would have first an empirical character, through which its actions, as appearances, would stand through and through in connection with other appearances in accordance with constant natural laws . . . [and] second . . . an intelligible character, through which it is indeed the cause of those actions as appearances, but which does not stand under any

conditions of sensibility [including universal causation] and is not itself [an empirical object]. (A539/B567)

This distinction between empirical and intelligible character makes it possible for Kant to defend the possibility of what he calls “transcendental freedom,”¹⁴ a power “of beginning a state from itself, the causality of which does not in turn stand under another cause determining it in . . . accordance with the law of nature” (A533/B561). We cannot rule out the possibility that humans, as things-in-themselves, have an intelligible character that is transcendently free in this sense. But this intelligible character can itself be the ground of an empirical character, and one who observes this empirical character will be able to trace empirical causes for any particular action.

The result is an initially shocking but ultimately quite plausible account of the relationship between freedom and natural necessity, one that distinguishes Kant’s account from the dominant accounts of freedom and causal necessity both in his day and our own. Many philosophers are “compatibilists,” who argue that freedom is compatible with causal determination. Generally, compatibilists define freedom as determination by internal, psychological causes rather than external ones, such that if an action follows from my choice, it is free, even if my choice is determined by external factors. Other philosophers defend incompatibilism, the view that freedom and natural determination conflict with one another. Such philosophers can be either “hard determinists” who believe that every event in the world, including every human choice and action, is causally determined by some set of prior conditions¹⁵ and therefore argue that there is no room for any meaningful kind of freedom, or “libertarians” who believe that (some) events in the world are determined by human choices that these choices are not in any sense determined by prior conditions.¹⁶ Kant’s position has aptly been called a “compatibility of compatibilism and [libertarian] incompatibilism” (Wood 1984: 74). Like incompatibilist libertarians, Kant defines freedom in a way that excludes prior causal determination of one’s choices, but like compatibilists, Kant believes that there is a way in

¹⁴ Kant’s terminology is somewhat misleading here because the freedom that Kant calls “transcendental” is really “transcendent.” It is not a condition of the possibility of experience nor evident from-within theoretical reasoning, but “transcends” any possible experience.

¹⁵ This is the dominant form of hard determinism among secular philosophers today. Other classic forms of determinism (or “predestination”) claim that every event is determined by Fate or by God.

¹⁶ Among libertarians, some (e.g. Kane 1996, 2002) emphasize natural indeterminacy according to which certain choices are undetermined natural events and others (e.g. vanInwagen 1986) emphasize “agent causation,” where events that are undetermined by prior conditions are explained as being the effects of agents who are responsible for them.

which one can assert both that something is freely caused and that something is the result of prior empirical causes. What makes Kant unique among contemporary theories is that he preserves a thoroughgoing causal necessity but at the same time an undetermined freedom. Kant's transcendental idealism allows him to see free things-in-themselves as grounds of the empirical world, while his empirical realism allows him to insist that within that empirical world, causation universally proceeds according to natural laws.¹⁷

The Critique of Pure Reason, where Kant develops this account of freedom, does not posit that such freedom is actual. This *Critique* focuses on what can be known a priori about the objective world, the world as it exists for human knowers. And what can be known about that world is that every event, including every human action, is causally determined by prior conditions in accordance with natural laws. Kant's discussion of freedom and causation is directed not only against those who argue for the impossibility of freedom (say, on the basis of Newtonian physics), but also against those who would claim that one can know that freedom exists in the world. In the *Critique of Pure Reason*, the "single thing we could accomplish and . . . our sole concern" is "to show that . . . nature at least does not conflict with causality through freedom" (A558/B586).

The theory of freedom that Kant lays out in the first *Critique* is thus presented as an option that metaphysics can neither establish nor rule out, an "extension" that "even if . . . empty, . . . we . . . can fill through practical data of reason" (Bxxi). And this sort of modesty shows an important positive aspect of Kant's limitations of metaphysics. After noting how his account of cognition precludes metaphysical proofs about things like God, humans' immortal souls, and freedom, Kant insists that while

a critique that limits . . . is, to be sure, to that extent negative, . . . because it simultaneously removes an obstacle that limits or even

¹⁷ There are two dominant ways that Kantians interpret this position. So-called "two-world" theorists read Kant as positing two metaphysically distinct "worlds," a noumenal world of things-in-themselves and a phenomenal world of appearances. The former includes humans insofar as we are free, the latter humans insofar as we are determined. And the former is the "ground" of the latter. Alternatively, so-called "two-standpoint" theorists claim that Kant posits only a single world that can be thought of in two different ways, as the sum of objects of possible experience or as a merely thinkable abstraction. When thinking of the world in the former way, freedom is precluded, but not when thinking of it in the latter way. Because morality requires thinking of ourselves as free (as we will see in the next section), the "merely thinkable" perspective gets content as a practical perspective from which we hold ourselves responsible. Thus insofar as human beings take an agent-standpoint on the world, we must view human beings as free. Insofar as we take a scientific-observer standpoint, we must see everything (including human beings) as causally determined.

threatens to wipe out the practical use of reason, this critique is also in fact of positive and very important utility, as soon as we have convinced ourselves that there is an absolutely necessary practical use of pure reason (the moral use), in which reason unavoidably extends itself beyond the boundaries of sensibility. (Bxxv)

Or, as he puts it much more succinctly later, “I had to deny knowledge in order to make room for [practical] faith” (Bxxx). Given his Copernican turn, Kant was able to use a transcendental anthropology of cognition to justify not only epistemic claims about the nature of human knowledge but even metaphysical claims about the nature of the empirical world. But precisely because such claims are limited to objects of possible experience, Kant makes room for non-empirical claims, if there is any non-cognitive access that human beings have to things-in-themselves. And Kant finds this non-cognitive access in another part of his transcendental anthropology, the transcendental analysis of volition wherein *morality* provides a non-cognitive role for reason in governing human life.

III. What ought I to do?

Kant's moral philosophy as transcendental anthropology of volition

From the question “What can I know?” Kant turns to the question, “What ought I to do?” While Kant’s transcendental analysis of cognition focused on human beings as human beings as free but finite knowers, Kant aims here to think about human beings as free but finite doers (or

agents).¹⁸ As in the case of cognition, Kant focuses on human actions “from-within,” and in particular explores both the norms governing human action and the conditions of possibility of being governed by those norms. Through laying out both the nature of action-guiding norms and the conditions of possibility of being bound by these, Kant offers insight – though not “knowledge” in the strict sense – into what human beings are in themselves. In particular, Kant’s moral philosophy completes the argument for human freedom by showing that such freedom is not only possible, but actual¹⁹, and by laying out “laws of freedom” that govern free human beings (G 4:387).²⁰

¹⁸ Kant’s *Groundwork* might seem specifically to avoid developing ethics as a subset of anthropology: “a law, if it is to hold morally, . . . must . . . hold not only for human beings, as if other rational beings did not have to heed it . . . [T]herefore the ground of obligation here must not be sought in the nature of the human being . . . but a priori simply in concepts of pure reason” (4:389, see too 4:410-12, 425). Kant is deeply opposed to thinking of morality as a subset of human biology or psychology, explicitly rejecting approaches to ethics that start with “conditions of human volition . . . drawn from psychology” (4:390-1). Thus *Groundwork* discounts what Kant calls “practical” or “moral anthropology” as merely a subsidiary part of ethics, one that “would deal only with the subjective conditions in human nature that hinder people or help them in fulfilling the laws of a metaphysics of morals” (G4:388, MS 6:217). The core of morality, Kant insists, must be “pure.” But this dismissal of anthropology at the core of morals is really only a dismissal of *empirical* anthropology at that core. As *Groundwork* progresses, the centrality of the human being as a free but finite chooser emerges clearly, as this section will show.

¹⁹ Both this “possibility” and “actuality” need to be interpreted carefully. Strictly speaking, the *Critique of Pure Reason* did not show that human freedom was really possible, only that the fact that human beings as they appear in the world are governed by natural laws does not necessarily preclude the possibility of humans being free “noumenally” or “in themselves.” And, as we will see, the “actuality” that is established in Kant’s moral philosophy is, metaphysically speaking, not an actuality in the objective world (since the “objective” world is a world of appearances), and epistemically speaking, not an actuality of which we can have “knowledge” in Kant’s strict sense.

²⁰ Whereas Kant’s Critical treatment of the question “What can I know?” is concentrated in his *Critique of Pure Reason*, his answer to “What should I do?” is scattered throughout several different texts. As early as 1765, Kant claimed to be working on a “Metaphysical First Principles of Practical Philosophy” (10:56) and in 1768 he assured a former student (Herder) that his “Metaphysics of Morals” would be done within a year (10:74). Later, he wrote that when he finished his “critique of pure reason” he would then develop his “metaphysics of nature and metaphysics of morals” (10:145). A *Metaphysics of Morals* would, in fact, have made a nice complement to the *Metaphysical Foundations of Natural Science*, and Kant’s *Critique of Pure Reason* even suggests such a parallel (A841/B869). In fact, however, Kant’s first Critical work addressing the question “What should I do?” was neither a Critique nor the promised *Metaphysics of Morals* but a mere *Groundwork for the Metaphysics of Morals* (1785), in which Kant aims at “nothing more than the search for and establishment of the supreme principle of morality” (4:392). This work thus addresses the question “What should I do?” only in the most general terms. Kant would subsequently write a *Critique of Practical Reason* (1788), which recapitulates the overall claims of the *Groundwork* but greatly elaborates the psychology and metaphysical implications that Kant sees as connected with this general moral theory. Not until shortly before his death would Kant finally publish his long promised *Metaphysics of Morals* (1797), in which he takes the general moral framework of the *Groundwork* and articulates a whole theory of the most important political and ethical obligations of human beings. Throughout this section, I aim to

Kant's argument for the actuality of freedom is based on the nature of moral obligation. For Kant, the from-within standpoint of volition – where one seeks to discern what one ought to do – has two important features relevant to human freedom. First, anyone who asks, in the broadest sense, what to do, “must regard itself as the author of its principles independently of alien influences; consequently, . . . as the will of a rational being it must be regarded of itself as free” (4:448). All choice happens “under the idea of freedom” (4:448) because the “power of choice . . . cannot be determined to action through any incentive *except so far as the human being has incorporated it into his maxim*” (6:24). This “Incorporation Thesis”²¹ claims that, from-within the standpoint of deliberation, all inclinations and incentives appear only as *candidate* reasons for action; one must “incorporate” them into one’s plans for action before they actually motivate. From-within, one sees this incorporation as something “free.”

For some contemporary Kantians, this analysis of the deliberative perspective from-within is sufficient to establish human freedom, but Kant worries that this argument does “not prove freedom as something *real*” but only as a necessary but possibly illusory “presupposition” (4:448-9). Thus Kant turns from the generic perspective of deliberation to the more specific stance of one asking the question, “what *ought I to do?*”, where “ought” is specifically *moral*. In the *Critique of Practical Reason*, Kant insists that the moral “ought” is ever-present within human practical deliberation: “the moral law, *of which we become immediately conscious (as soon as we draw up maxims of the will for ourselves)*, . . . offers itself to us and . . . reason presents it as a determining ground not to be outweighed by any sensible conditions and indeed quite independent of them” (5:29-30, emphasis added). In the process of devising and considering principles to act on, we become “immediately conscious” of a “fact of reason” (5:29), the fact that we are bound by a moral law, one that commands obedience regardless of other incentives. This fact “forces itself upon us of itself as a synthetic a priori proposition that is not based on any intuition” and not dependent upon reasoning to obligation from any “antecedent data of reason” (5:31).

From this fact of reason, Kant aims to establish that human beings are (transcendentally) free by showing that “a [transcendentally] free will

offer a synchronic account of Kant's moral philosophy drawing from all of his “Critical” works (starting with the *Groundwork*). In several cases, Kant's positions on important issues in moral philosophy shifted between these works, most notably with respect to his arguments for freedom and morality, but this section ignores those changes. For discussion of some of them, see Ameriks 2003: 161-192.

²¹The claim was given this label by Henry Allison. See Allison 1990: 5, 40.

and a will under moral laws are one and the same” (4:447, cf. 4:450, 5:28-9).²² In order to establish this mutual implication, Kant draws on “common rational moral cognition” to “search for and establishment of the supreme principle of morality” (4:392). What could the supreme principle of morality be? To answer this question, Kant focuses on two (related) features of the moral ought, its independence from inclination and its universality. Moral reasons are distinguished from other sorts of reasons in that they are not tied to things that one happens to find oneself wanting. When one decides that one “should” buy gasoline for one’s car, one does so only because one thinks that such an activity will be conducive to ends that one happens to have. One can always decide to forgo those ends, and then one need not buy gasoline. But when one decides that one “ought” to refrain from falsely accusing an innocent adversary or “ought” to help a stranger in immediate pressing need, one does not see these decisions as optional in the same way. It does not matter whether the false accusation fits with other goals that one has, nor whether one cares about the stranger. *Moral* obligations do not depend upon our inclinations. Kant puts this point in terms of a distinction between what he calls “hypothetical imperatives,” which are commands that one has to obey *if* one wants to achieve some particular end, and “categorical imperatives,” which are (moral) commands that one simply *has to obey no matter what* (no “if”-clause). Relatedly, Kant argues that the moral law is *universal*: “everyone must grant that a law, if it is to hold morally . . . must carry with it absolute necessity,” going so far as to say that “the command ‘thou shalt not lie’ does not hold only for human beings, as if other rational being did not have to heed it” (4:389). Kant’s point here is not that everyone ought always act in the same way. Someone who cannot swim need not jump into a river to save a drowning child, and someone with remarkable artistic talents may have an obligation to cultivate them that others would not have. The point, rather, is that morality itself is universal, in that when one becomes immediately conscious of obligation in general, one is conscious of it as a law that binds everyone (even if it binds different people in different ways).²³ Another person *who is relevantly similar* to me (able to swim, or possessed of similar talents) will have the same obligations. Unlike

²² Given this “circle” (4:450) or reciprocal implication (5:29), establishing either that one is under moral laws or that one is free is sufficient to establish both. Although Kant initially (in the *Groundwork*) tries to establish the validity of the moral law by first offering an independent and quasi-theoretical argument for human freedom, his mature position is that we must first know the validity of the moral law, which is given as a fact of reason and cannot be deduced from any prior knowledge, and only from our knowledge of this fact of reason can we know that we are free.

²³ This is most evident in applications of the moral law that are least situation-dependent, such as “do not falsely promise,” since in this case, the particular moral requirement is universal as well.

inclinations, morality is not something that one can pick and choose. It *obligates* everyone.

Given these characteristics of morality, one might think that it would be impossible to derive a fundamental formula of morality. If all that we know about morality in general is that it can derive from neither particular inclinations nor contingent features of ourselves, then there seems to be nothing left from which to get a “principle” of morality at all. But in fact, Kant argues that the limitations on the content of the moral law actually give rise to a “formula” that encapsulates the fundamental principle of morality.

When I think of a hypothetical imperative in general I do not know beforehand what it will contain; I do not know this until I am given the condition [i.e., the end to be promoted]. But when I think of a categorical imperative I know at once what it contains. For, since the imperative contains, besides the law, only the necessity that the maxim be in conformity with the law, while the law contains no condition to which it would be limited, nothing is left with which the maxim of action is to conform but the universality of a law as such; and this conformity alone is what the imperative properly represents as necessary.

There is, therefore, only a single categorical imperative and it is this: act only in accordance with that maxim through which you can at the same time will that it become a universal law. (4:420-1)

The moral law of which I am immediately conscious within deliberation is a law that commands me to act only in such a way that the bases for my actions – my “maxims” – could be bases for the actions of everyone. What is universally commanded to all is the practice of acting in a way that could be universal for all.

Kant goes on to redefine this categorical imperative based on a particular feature of human willing: human beings not only follow various practical laws, but also act for the sake of ends (4:427). Now the moral law is not determined *by* any particular (contingent) ends, but it does determine a necessary end, “something the existence of which in itself has an absolute worth, something which as an end in itself could be a ground of determinate laws” (4:428). And Kant finds just such an end in “the human being” (4:428). This gives Kant a new way of describing the categorical imperative:²⁴ “So act that you use humanity,

²⁴ For discussion of the relationship between the different formulations of the categorical imperative, see Korsgaard 1996a and Wood 1999.

whether in your own person or in the person of any other, always at the same time as an end, never merely as a means” (4:429, see too 6:462). This new formulation of the moral law puts human beings at the center of morals, not only in that the moral law is derived from a transcendental anthropology of volition, but also in that the ultimate end of morality, that which must at all times be respected, is nothing more (nor less) than the human being.²⁵

Finally, Kant adds another formulation of the categorical imperative that further enriches his transcendental anthropology and paves the way for his defense of human freedom. Kant insists that “the human being is . . . subject only to laws given by himself . . . and is bound only to act in conformity with his own will” (4:432). Kant describes this independence from external laws as “autonomy” and points out that this autonomy does not imply lawlessness, but rather that one is subject always only to one’s own laws. This may seem to be merely a recapitulation of the Incorporation Thesis, but Kant’s point here is more specific. If the moral law is to be truly universal and independent of our inclinations, then it cannot be derived from anything external to our will itself. Any *external* command would need to appeal to us for some reason, either because we feel inclined to obey it (in which case it is not truly moral) or because we *ought* to obey it (in which case its authority derives from rather than grounds morality). For moral laws to be truly one’s own rather than merely results of outside influences manipulating our contingent desires, autonomous lawgiving must proceed by means of laws that have no basis other than our own wills. But laws determined solely by our wills are categorical. So, for Kant, “autonomy of the will [is] the supreme principle of morality” (4: 440).²⁶

At this point, Kant has *nearly* proven that human beings are transcendently free. The principle of morality is a principle of autonomy, or *self*-governance. But to make the stronger claim that this “autonomy” is identical with transcendental *freedom*, Kant goes further. He offers a quasi-geometric proof starting with the nature of moral obligation and deriving the necessity of transcendental freedom. He poses the following problem:

²⁵ Precisely what Kant means by humanity here is hotly contested. See Dean 2006 and Frierson 2007.

²⁶ Kant takes this argument one step further. The moral law is “the will of every rational being as a willing giving universal law” (4: 431). But this universal sort of autonomy is tied, for Kant, to the fact that “a rational being belongs as a member to a kingdom of ends when he gives universal laws in it but is also himself subject to these laws” (4:433). Respect for humanity in oneself and others involves submitting oneself only to laws that one legislates for oneself, but also legislating laws for oneself that one also, at the same time, legislates for all rational agents. Thus human beings are not only autonomous, but autonomous members of a kingdom of ends.

Supposing that the mere lawgiving form of maxims is the only sufficient determining ground of a will: to find the constitution of a will that is determinable by it alone. (5:28)

That is, Kant considers what sort of will could be determined by a moral law that dictates only the “form” that one’s maxims must take and says only that such maxims must be universalizable, without saying anything about the “matter” of those maxims, that is, what sorts of goals one should aim for in one’s actions. Kant argues,

Since the mere form of a law . . . is not an object of the senses and consequently does not belong among appearances, . . . this form as the determining ground of the will is distinct from all determining grounds of events in nature . . . , [so] a will [determined by this ground] must be thought as altogether independent of the law of causality. (5:28-9)

From-within the standpoint of deliberation, when one considers whether or not to act on the basis of the moral law, one precisely sees this law as a law that offers nothing to one’s natural inclinations. There is, in that sense, no “natural” basis for acting in accordance with it. When one chooses to act on an ordinary inclination – say, deciding to eat an appetizing cookie – one can see oneself as “giving in” to the flow of natural causes. But because its demands are fundamentally formal, the moral law is not the sort of thing that one can merely “give in” to. It “presents it[self] as a determining ground not to be outweighed by any sensible conditions and indeed quite independent of them” (5:29-30). Thus the only will that can be truly bound by the moral law is a will that is free from sensible (that is, empirical) conditions. But freedom from determination by empirical conditions is precisely what transcendental freedom is, so a will under the moral law is a transcendently free will. Kant decries any traditional form of compatibilism as “wretched subterfuge” (5:96): “Psychological or comparative” freedom, where “free” just means that “actions are caused from within,” is “nothing better than the freedom of a turnspit” (5:96-7).

To his abstract argument and violent polemics, Kant adds a more intuitive thought-experiment to show that when we reflect on actions from-within, in terms of what we take ourselves to be capable of, even apparently irresistible temptations are eminently resistible:

Suppose someone asserts of his lustful inclination that, when the desired object and the opportunity are present, it is quite irresistible to him; ask him whether, if a gallows were erected in front of the

house where he finds the opportunity and he would be hanged immediately after gratifying his lust, he would not then control the inclination. One need not conjecture very long what he would reply. But ask him whether, if his prince demanded, on pain of the same immediate execution, that he give false testimony against an honorable man whom the prince would like to destroy under a plausible pretext, he would consider it possible to overcome his love of life, however great it might be. He would perhaps not venture to assert whether he would do it or not, but he must admit without hesitation that it would be possible for him. (5:30).

The first part of this thought-experiment shows only that human beings are capable of overcoming particular sensuous desires (such as lust) when the fulfillment of these threatens more important sensuous desires (such as love of life). But the second part shows that human beings recognize in themselves an ability to overcome even love of life for the sake of the moral law. If our love of life can motivate us to overcome our everyday sensuous desires, and our respect for the moral law can motivate us to overcome even our love of life, then there is no temptation that we are unable to overcome for the sake of the moral law.

Importantly, Kant is not denying that, from-without, our actions have empirical causes. Even consciousness of the moral law appears as an empirical cause in a chain of mental events that gives rise to a volition to act in accordance with it. But from-within volition, we become aware of a sense of responsibility the condition of possibility of which is the transcendental freedom that, properly understood, Kant's *Critique of Pure Reason* showed to be compatible with causal-determinist explanations from-without. Through examining the moral law present from-within in human volition, Kant shows that humans are transcendently free and thus "fills the vacant space" (5:49) left open by his theoretical philosophy. But Kant does more. By specifying the most fundamental principle of morality, Kant fills this vacant place "with a determinate law of . . . an intelligible world . . . , namely the moral law" (5:49). That is, Kant shows not only that human beings are free, but also that human freedom is not lawless and arbitrary but a law-governed capacity to be moral.²⁷

²⁷ Just as Kant showed that the moral law implies freedom, he also shows that freedom implies morality. The point here is that if we think about what a law of freedom must be, we know that we cannot derive such a law from anything about what our natural motives and interests happen to be, since these are all determined by laws of nature. But we also know that any law of freedom, precisely because it is not based on contingent empirical details, will be the same for all rational agents. And so, for Kant, the "content" of the moral law is simply putting these formal criteria of morality into the form of a principle governing the

Humans' sense of moral obligation, properly understood, provides evidence of freedom and also gives rise to a specific principle of morality.

Kant's arguments for transcendental freedom as central to human nature are hardly beyond controversy,²⁸ and the rest of this section focuses on two key problems that arise for the Kantian account of freedom and morality offered so far. The first problem is this: if human beings are *really free* only insofar as we submit to the moral law, Kant seems unable to account for the possibility of human beings ever being responsible for doing what is morally *wrong* (see, e.g. Sidgwick 1901 and Reinhold 2006). If the moral law is the law of freedom, then whenever human beings act contrarily to the moral law, they must not really be free. But freedom is a condition of possibility of moral responsibility, so whenever human beings act wrongly, they are seemingly not morally responsible for their actions. Kant does claim that human beings can be held responsible for acting badly,²⁹ but how can he do this? Second, given that Kant's transcendental anthropology of desire is intended not merely to lay out the conditions of possibility of moral responsibility but also to clarify precisely what, from the standpoint of deliberation, humans find themselves obligated to do. But if Kant's moral philosophy is supposed to answer the question, "What ought I to do?," the mere formula of universal law (FUL) – "*act only in accordance with that maxim through which you can at the same time will that it become a universal law*" (4:421) – seems too abstract to provide real guidance for action.

Regarding the first problem, as important as freedom is to his transcendental anthropology, Kant recognizes that human choosers are not *merely* morally free beings. Even from-within the perspective of human volition, we find ourselves to be *both* free beings subject to the moral law *and* members of the sensible world, subject to empirically-informed desires and inclinations. Even Kant's "pure moral philosophy" articulates what morality means for beings like us, who participate in

will: whatever you do, make it the kind of thing that could be done from freedom, that is, that could be a law for anyone.

²⁸ Among the problems with the argument that I do not discuss in detail here: (1) Kant's appeal to a "fact of reason" makes the argument ineffective against true moral skeptics. (2) Compatibilist freedom is, today, generally considered in much better shape than Kant's polemics allow, and even many of Kant's supporters think that his moral theory can do without a strongly metaphysical account of transcendental freedom. (3) Kant's account of the relation between freedom and morality poses specific problems for making sense of how finite free beings like ourselves can be bound by the moral law; if humans as empirical entities in the world are bound by natural, causal laws, how would we ever recognize the free submission of a human being to morality? Later chapters discuss compatibilism (chapter eight) and the empirical expression of morally good action (chapter two).

²⁹ We examine this claim in detail in chapter three.

both an intelligible world governed by laws of freedom and a sensible world governed by laws of nature.³⁰ Because of our sensible nature, human beings have various natural inclinations that can conflict with the demands of the moral law in particular circumstances.³¹ It is because we have such non-moral inclinations that morality takes the form, for us, of “duties” and “imperatives,” commands that we *ought* to obey rather than a moral law that we simply *do* obey (4:413).

In that context, Kant distinguishes between “positive” and “negative” freedom.³² Negative freedom is a “property in us . . . of not being necessitated to act through any sensible determining grounds” (6:226, cf. 4:446), while positive freedom is the property of acting through a non-sensible determining ground, the moral law (G 4:446-7, MM 6:213-14). The former, negative freedom, is necessary in order to hold human beings morally *responsible*, while the latter, positive freedom, is what constitutes the full-blown autonomy of a morally *good* agent. Too sharp a line between these two sorts of freedom would undermine moral autonomy. If negative freedom is not identical to positive freedom, why should we see the moral law as arising from the free choice that makes us morally responsible? Why see the moral law as any less alien to our (negatively) free selves than natural causes? In response, Kant argues for an intrinsic link between negative and positive freedom. Insofar as negative freedom is a freedom from having one’s actions governed by anything external to oneself, the only way to remain free is to make one’s law the law of freedom, the categorical imperative. As one commentator has put it, “by making the [categorical imperative] its principle, the free will retains the position of [freedom]” (Korsgaard 1996a:166, cf. 6:227). By contrast, “the free will that puts inclination above morality sacrifices its freedom for nothing” (Korsgaard 1996a:167). Not only does (negative) freedom make morality possible by freeing us from inclination, but the moral law itself is identical with the internal constitutive standards of freedom. Because the moral law specifies nothing other than the

³⁰ Thus while Kant’s *Groundwork* begins by discussing the “good will” in general, which belongs to God and can also belong to human agents, Kant quickly specifies the nature of this will in such a way that it applies more particularly to human wills: “we shall set before ourselves the concept of . . . a good will though under certain subjective limitations and hindrances” (4:397).

³¹ In his *Groundwork* and *Critique of Practical Reason*, Kant names the comprehensive satisfaction of these interests “happiness,” and he explains that human beings are always tempted by a “principle of self-love” rooted in our sensible nature, a principle that ought to be – but isn’t always in fact – restricted by the moral law rooted in our free, rational nature. The scope of natural inclinations is very broad, such that even such things as a natural philanthropy that “finds an inner satisfaction in spreading joy” (4:398) is a merely natural inclination. “Self-love” has different forms.

³² This distinction is often linked with Kant’s distinction between the “will” (German *Wille*) and “choice” (German *Willkühr*). The will, which Kant identifies with “practical reason,” has positive freedom. “Choice” is the human capacity by virtue of which we have negative freedom.

condition of freedom, choice based on any other principle limits rather than reaffirms freedom. Human beings are always negatively free, in that we *need not* let our actions be determined by forces external to us, but we are not always positively free, since we often relinquish autonomy in the face of temptation.

The second problem, as Hegel classically put it, accuses Kant's categorical imperative of being an "empty formalism," an "abstract universality, whose determination is . . . without content" (Hegel 1991:162). Precisely because this "specific principle of morality" is purely formal, it gives only the most abstract account of what is required of human beings. In order to complete his account of the norms that ought to govern human volition, Kant must deliver a more complete framework of normative constraints on human volition. Hegel and others have argued that this will require "bringing in material *from outside* [to] arrive at *particular* duties [because] it is impossible to make the transition to . . . particular duties . . . from the determination of duty as *absence of contradiction . . . with itself*." ³³ Kant's abstract moral law seems insufficient to provide moral content from within.

Kant has a two-fold way of dealing with Hegel's empty formalism objection. First, even if the categorical imperative is, in itself, an empty formalism, it can still be action-guiding in an important way. Hegel suggests that the categorical imperative would only prohibit the stealing of property, for example, if one has independent bases for thinking that property rights are good. But insofar as one tests *maxims* for action, one can evaluate those maxims based on the values implicit within them, without ascribing any independent normative weight to those values. Thus the thief who acts on the maxim, "I will steal my neighbor's car in order to have it for myself," commits herself to the value of private property by virtue of her end (having it *for herself*), and thus her maxim conflicts with the categorical imperative. The native Hawaiians who acted on the maxim "We will take the iron nails out of the bottom of that ship in order to make spears" did not violate the categorical imperative (at least not directly³⁴) because their maxim implied a commitment to the value of spears over ships, but not any direct commitment to institutions of private property. And to this extent, at least, even if the categorical imperative is insufficient for evaluating the moral status of *actions*, it does seem to be an important way of picking out certain *maxims* that, because they require making an exception of oneself, are morally wrong.

³³ Hegel, *Philosophy of Right*, ed. Allen Wood, Cambridge 1991, § 135, p. 162.

³⁴ Kant does argue that a commitment to private property is implied by any external use of one's freedom, so the native Hawaiian's maxims imply, for Kant, an indirect commitment to private property.

Second, Kant's emphasis in *Groundwork* on pure moral philosophy is explicitly only a *foundation* for a complete "metaphysics of morals." Just as the empirical concept of matter is needed to move from the metaphysics of the *Critique of Pure Reason* to the basic principles of physics, empirical attributes of human beings are needed to move from a general principle of morals to specific moral duties. As the particular kinds of embodied, finite agents that human beings are, we have specific talents, needs, strengths and limitations that give rise to specific duties. The normative force of these duties comes from their connection to the fundamental moral principle by virtue of which human beings are rational, free, autonomous agents. But the specific content comes from the way that we must act in order for our empirically discoverable needs and desires to be satisfied through acting on maxims that conform to that fundamental moral principle (6:217).

The result, when Kant turns to his *Metaphysics of Morals*, is a detailed account of human obligations in the face of our finite natures. First, Kant points out that for *embodied* free beings, freedom manifests itself through actions in an empirical world. From this empirical claim, Kant arrives at his "universal principle of right":

Any action is *right* if it can coexist with everyone's freedom in accordance with a universal law, or if on its maxim the freedom of choice of each can coexist with everyone's freedom in accordance with a universal law. If then my action or my condition generally can coexist with the freedom of everyone else in accordance with universal law, whoever hinders me in it does me *wrong*. (6:230)

Interfering with another's activity or condition is wrong, unless that other's activity or condition is itself wrong. This general principle of non-interference involves translating Kant's general formula of universal law specifically in terms of human *actions* rather than maxims, and it grounds what becomes Kant's political theory. Political "coercion that is opposed to [wrong actions]" is justified as a "hindrance to a hindrance of freedom" (6:231), and Kant specifies political rights based on increasingly detailed empirical descriptions of human beings, ranging from the need for private property rights (since human beings depend upon external goods to exercise their freedom) to parental rights and duties (because of the vulnerability of children). The result is a detailed *Doctrine of Right*, which lays out actions that not only are required or prohibited by the moral law but also ought to be required or prohibited by enforceable political laws.

Second, given the structure of human volition, which always acts for the sake of ends, there must be “ends that are also duties” (6:385). Because humans are dependent upon their own abilities and those of others to accomplish any ends at all, we have obligations both to promote our own abilities (for the sake of accomplishing whatever future goals we might find ourselves to have) and to use our abilities to help others accomplish their ends (since we, like them, are dependent upon others’ assistance).³⁵ From the individual abilities that we can improve in our own case, Kant derives duties prohibiting such things as suicide and lying and requiring such things as self-examination and the deliberate cultivation of “powers of spirit, mind, and body” (6:444). From our obligation to make the happiness of others an end, he derives virtues such as beneficence and gratitude and vices such as arrogance, ridicule, and contempt. Kant goes on to discuss friendship as “the most intimate union of love with respect” (6:469f.) and to mention duties that arise from specific, empirically-given “differences in rank, age, sex, health, prosperity or poverty, and so forth” (6:469).

In the end, Kant’s transcendental anthropology of desire offers a detailed answer to the question “What ought I to do?” and in the process further expands on the conception of human beings as free and finite beings that Kant began in the *Critique of Pure Reason*. Not only are we free and finite *doers* as well as *knowers*, but because transcendental freedom is a condition of possibility of the moral obligation under which we find ourselves from-within the standpoint of choice, we can justifiably believe that humans are transcendently free things-in-themselves, even though we can never strictly “know” this. Kant’s promise in the first *Critique* that he would “deny knowledge to make room for faith” (B xxx) is fulfilled in his moral philosophy. In the process, the “faith” for which he held out hope in the *Critique* is shown to be not a *blind* faith, but a solid conviction grounded in rational arguments based on the conditions of possibility of moral responsibility. Kant calls such morally-grounded beliefs “postulates” and says,

³⁵ We do not have a “duty” to promote our own happiness since this is something “everyone already wants unavoidably” (6:386), and Kant (confusingly) argues that we do not have a duty to promote others “perfection” since such perfection “consists just in this: that *he himself* is able to set his end in accordance with his own concepts of duty” (6:386). This is confusing, since Kant generally uses “perfection” to refer to nonmoral as well as moral perfections, and it is not clear that, for instance, developing a better memory requires developing that memory *for oneself*. But this confusion is relatively benign; insofar as perfections are non-moral abilities to achieve ends, the obligation to promote others’ happiness includes an obligation to promote their perfections. Insofar as “perfection” is specifically moral and thus not necessarily linked to (earthly) happiness, his reason for excluding it applies.

All of them proceed from the principle of morality, . . . [which] requires these necessary conditions [such as freedom] for the observance of its precept. These postulates are not theoretical dogmas but *presuppositions* having a necessarily practical reference and thus, although they do not indeed extend speculative cognition, they give objective reality to the ideas of speculative reason . . . (by means of their reference to what is practical) and justify holding concepts even the possibility of which it could not otherwise presume to affirm. (5:132)

Kant even claims that this morally-grounded faith has a sort of “primacy” (5:119) over knowledge one has through empirical cognition; one’s theoretical reasoning about the world must “accept” the postulates “as soon as these propositions belong inseparably to the practical interest of pure reason” (5:121). The transcendental anthropology of desire fills in the otherwise “vacant space” for freedom left by human cognition (5:49, 103).

Before closing this section, it is worth attending to one further, dramatic aspect of Kant’s transcendental anthropology of desire. In the *Critique of Pure Reason*, Kant had highlighted *three* traditional problems of metaphysics that would be stricken from the realm of knowledge – “God, freedom, and immortality” (Bxxx, A3/B7) – and in his transcendental anthropology of volition, Kant comes back not only to freedom but to the issues of God and immortality as well. As in the case of freedom (though to a different degree), Kant argues that belief in God and immortality are practically necessary. Neither God nor immortality are conditions of the possibility of moral responsibility *per se*, but when Kant considers what the ultimate goal of a virtuous agent must be, he argues that while the “supreme end” will be virtue alone, the “complete” end – that end from which nothing good is absent – must include both virtue and “happiness distributed in . . . proportion to morality” (5: 110). The reason for this is not that the agent merely *wants* happiness, but that “an impartial reason” could not deliberately choose a world within which some beings “need happiness, [and are] worthy of it, and yet [do] not participate in it” (5:110). Insofar as virtuous agents seek this highest good, they must believe in whatever is necessary in order for their activity to reasonably be held to contribute to this highest good. For Kant, immortality is necessary because virtue can never be fully realized in one’s finite life but only in endless progress (5:122). God is necessary in order to ensure that happiness is doled out in proportion to virtue (5:124ff.). Only by believing in both God and immortality can our efforts towards virtue be reasonably taken to be efforts towards the complete highest good.

Kant's arguments for God and immortality are more complicated than I have suggested here, and their validity is widely disputed. For the purposes of understanding Kant's conception of human beings, the details of these arguments are less important than the overall implication of Kant's approach. Just as Kant in the *Metaphysics of Natural Science* makes Newtonian physics a subset of a transcendental anthropology of cognition, he here makes traditional *theology* a subset of a transcendental anthropology of volition. By the end of his transcendental anthropology of cognition, Kant had shown that the a priori structure of human cognition establishes (among other things) our ability to know an empirical world as consisting of substances in causal relationships with one another, and he offers a priori foundations for natural science. Having added a transcendental anthropology of volition, Kant has laid out the a priori laws governing the realm of free human agents and defended even God's existence as part of a philosophical *anthropology*.

IV. What may I hope?

The Critique of Judgment as transcendental anthropology of feeling

Given the results of the previous two sections, Kant's transcendental anthropology might seem complete. Human beings are free, finite knowers and doers, governed within each realm by a priori laws that we give ourselves. We exist as both fully free things-in-ourselves and finite, embodied appearances in the empirical world. Within the empirical world, we see ourselves and everything else as governed by natural laws. As free, we are governed by moral laws. Kant's first and second questions – about knowledge and obligation – have been answered, and the question “What may I hope?” seems answered by Kant's practical postulates of God and immortality.³⁶ Nonetheless, shortly after finishing his *Critique of Practical Reason*, Kant set to work on a third *Critique*, which would eventually become the *Critique of Judgment* and would provide the a priori laws of feeling that could complete his transcendental anthropology. By this time, Kant had made

³⁶ For a long time, in fact, Kant took his transcendental anthropologies of cognition and desire to complete his transcendental philosophy as whole and specifically “thought it impossible” to find a priori principles for the faculty of feeling (A21), and Kant's *Critique of Pure Reason* seemed to include analyses of “the logical functions of the understanding in judgments” (A70/B95) and a “transcendental doctrine of the power of judgment” (A137/B176), so a specific critique of judgment seemed out of place.

three realizations that required a rethinking of the nature of this anthropology.

First, Kant came to see that his earlier rejection of feeling as a faculty capable of transcendental investigation was mistaken. Kant had rightly seen that the pleasures that humans take in what is merely “agreeable” – food, sex, reputation, baseball – are empirically rooted and thus incapable of a priori investigation. But as he continued to teach and study aesthetics, he came to see that judgments that something is beautiful or sublime are at once subjective because rooted in feeling and taken to be universal and normative; to claim that something is beautiful is to claim that *all* others *should* find it beautiful. Normativity, perhaps even of an a priori sort, is applicable to aesthetic feeling. Second, Kant recognized that his account of the cognition of nature was incomplete in its application to the empirical world. His *Critique of Pure Reason* ensured that the world would conform to certain general structures of human cognition, but it provided no assurance that humans would be able to expand the scope of their knowledge in any systematic way. Finally, Kant’s moral philosophy was incomplete in its application to the empirical world. The *Critique of Practical Reason* provided an a priori argument to show that the end human beings are obligated to promote – the highest good – is possible, but it provided no basis for this possibility in the observable order of nature. Kant’s attempt to “deny knowledge in order to make room for belief” was insufficient to explain *how* nature and freedom relate to each other. He needed a *Critique of Judgment* to provide a “mediating concept between the concepts of nature and the concept of freedom, which makes possible the transition from the purely theoretical to the purely practical, from lawfulness in accordance with the former to the final end in accordance with the latter” (5:196, cf. 5:176). Kant came to see his accounts of cognition and of volition as insufficient in their application to the empirical world, and he lost his early pessimism about the possibility of a transcendental (a priori, normative, from-within) analysis of feeling. These realizations led Kant to complete his transcendental anthropology with a *Critique of Judgment* that would investigate the faculty of feeling through the power of judgment.

The general structure of the *Critique of Judgment* can seem perplexing, since it is divided into two halves that seem, at first blush, unrelated.³⁷ The first half – a “Critique of Aesthetic Judgment” – explores conditions of possibility of making justified aesthetic judgments about

³⁷ For a defense of the claim that they are not significantly related, see MacFarland 1974. For a detailed reading of the *Critique of Judgment* as a coherent whole, see Zuckert 2007.

beauty or sublimity. The second half – a “Critique of Teleological Judgment” – lays out Kant’s philosophy of biology, within which Kant argues that for the study of living things, one must make use of teleological principles in addition to the laws of mechanical causation defended in his *Critique of Pure Reason* and *Metaphysics Foundations of Natural Science*. Moreover, while the “Critique of Aesthetic Judgment” focuses on laying out an a priori principle that governs the faculty of feeling, there is no direct reference to feeling in the “Critique of Teleological Judgment.” And while the “Critique of Teleological Judgment,” especially with its discussion of the ultimate and final ends of nature, provides a transcendently-grounded framework for answering the question “What may I hope?,” the analysis of the beautiful and sublime seems irrelevant to answering that question. All of this can make it seem that however helpful this book might be in other respects, it cannot provide a unified transcendental anthropology of feeling that would complete Kant’s philosophy.

In fact, however, the book as a whole is unified by the principle of the purposiveness of nature, an a priori principle of judgment that provides a basis for universal norms governing feeling. Purposiveness emerges as an a priori principle of judgment in the context of Kant’s transcendental anthropology of aesthetic feeling, or taste (20:244). But this a priori principle can be applied more widely than *merely* to govern aesthetic feeling:

once the capacity of the power of judgment to institute a priori principles for itself is granted, then it is also necessary to determine the scope of this capacity, and for this completeness in critique it is required that its aesthetic faculty be recognized as contained in one faculty together with the teleological and as resting on the same principle, for the teleological judgment about things in nature also belongs, just as much as the aesthetic, to the reflecting power of judgment. (20:244)

Moreover, once purposiveness is established as an a priori principle, Kant can use it to address the insufficiencies of his transcendental accounts of cognition and volition. What starts in a transcendental anthropology of (aesthetic) feeling becomes the unifying principle of Kant’s transcendental anthropology as a whole and the basis for answering the final question of Kant’s philosophy: “What may I hope?”³⁸

³⁸ Ultimately, answering this question depends upon Kant’s philosophy of religion and history as well, but the *Critique of Judgment* provides at least a partial basis for this answer.

Before unpacking the details of this account, we should address the question of whether a transcendental anthropology of feeling is even appropriate. Recall that *transcendental* anthropology has at least three distinctive features: it is *a priori*, investigates humans from-within, and emphasizes normative constraints. Human *feelings* seem ill-suited to any of these sorts of analyses. Of all aspects of human life, feelings seem to be the most empirically contingent. And even though we can introspectively examine our feelings, there does not seem to be the sort of “from-within” relationship to feeling that we have with cognition and desire. Whereas we actively think and choose, feelings seem to be things that just happen to us. This point is tied to the final one; Kant’s “from-within” perspective is fundamentally normative, not a matter of how things seem to us but a matter of how we govern ourselves in thought or choice. And normativity does not seem appropriate to feeling; it is at least a bit odd to say that a person *felt* wrongly. And even if there is *some* sort of normativity governing feelings, it does not seem *a priori*. It seems that any anthropology of feeling should be empirical, a matter of introspection into what one does feel in various circumstances and observation of the feelings of others, with perhaps some prudentially-normative guidelines based on what makes for a good human life or some moral restrictions ultimately traceable to the moral principle at the heart of Kant’s transcendental anthropology of volition.

Kant raises many of these concerns himself. He points out that while there are “empirically knowable” connections between objects and natural feelings of pleasure that give rise to desires for those objects, such connections are “not grounded in any principle *a priori*” and thus do not provide suitable material for a transcendental anthropology of feeling (20:206, cf. A21). (Kant calls the objects of these pleasures “agreeable.”) Other objects might give rise to pleasure because they are useful in some way, and one takes pleasure in their suitability to some end. Such objects please because they are “good-for” something and their pleasure will be based in empirically knowable connections between those objects and the ends for which they are good. There is also respect for the moral law, which is both a feeling and required *a priori*, but it is required only by virtue of its connection with volition. The necessity of respect does not require a “special . . . critique of the feeling of pleasure and displeasure” but can be subsumed under a transcendental anthropology of volition (as Kant does in the *Critique of Practical Reason*). In fact, pleasure in *both* the agreeable *and* the good (whether useful or moral) can be explained by reference to the faculty of desire (or volition). Agreeable objects are the goals of hypothetical imperatives; the gratification we find in them “arouses inclination” (5:207). Useful objects

are the necessary or helpful means to some given ends, so they provide satisfaction “only as a means” (5:207). And the morally good is the object of the categorical imperative; we feel satisfaction in the morally good because of its connection to volition.

But Kant suggests that some pleasures are due to neither agreeableness nor goodness in their objects. These pleasures, for Kant, are judgments of “taste” or of “aesthetic pleasure” and have for their objects things that are “beautiful” or “sublime.” Kant’s primary focus is on the beautiful, and he structures his transcendental analysis of beauty around several key claims about how pleasure in the beautiful presents itself to us from-within: it is *disinterested* (5:204-211), it is non-conceptually *universal* and *necessary* (5:211-219, 5:235-40), and it presents its object as *purposive without a purpose* (5:219-235). For these sorts of pleasures, Kant argues, an a priori principle is both needed and available.

Kant’s first claim – that aesthetic pleasure is disinterested – merely emphasizes that beautiful objects are neither agreeable nor good and thus cause pleasure without connection to “interest” (that is, without arousing volition). To judge “whether something is beautiful” we consider it within the standpoint of “mere contemplation,” and from that standpoint, “one only wants to know whether the mere representation of the object is accompanied by satisfaction in me, however indifferent I might be with regard to the existence of the object” (5:204-5). For example, if I want to know whether “the palace that I see before me [is] beautiful,” I need know only whether the mere contemplation of it brings pleasure, even if, “were I to find myself on an uninhabited island . . . and could conjure up such a magnificent structure through my mere wish, I would not even take the trouble of doing so” (5:204-5, cf. 29:877-8).

The second claim brings up central dilemma that drives Kant’s analysis, the “reason why judgments of taste are subject to a critique with regard to their possibility” (5:191). Aesthetic judgments involve “not an empirical concept but rather a feeling of pleasure . . . which . . . is nevertheless to be expected of everyone” (5:191). For Kant, when one judges that some object is “beautiful” or “sublime,” one does not ascribe to that object anything that can be expressed in general concepts; one claims only that the object gives rise to a certain kind of pleasure. But one nonetheless takes one’s judgment to be “correct,” that is, one takes it that the object *should* give rise to that pleasure. And this is not merely a subjective claim; one takes it that the object should give rise to pleasure *in anyone*. Moreover, like other key claims in Kant’s transcendental anthropology, this insistence that human beings take pleasure in the

object is not an empirical-psychological claim; one does not claim that all others *will* or *do* feel this pleasure, but rather than that they *should*. Aesthetic judgments present a normativity that is reducible neither to epistemic norms (since epistemic norms apply to the formation and application of concepts) nor moral-practical (both because aesthetics is disinterested and because practical norms require an appeal to concepts). Because aesthetic judgments, seen from-within, are governed by a sort of normativity that is irreducible to epistemic and moral normativity, a separate transcendental analysis of the conditions of possibility of this normativity is required.

The final key claim about pleasure in the beautiful provides Kant's solution to his central dilemma, but it is also the most confusing of Kant's claims about the beautiful. Beautiful objects incite pleasure because they are "purposive without a purpose" (see 5:220). What does that mean? And why would the purposiveness of beautiful objects be a ground for universal pleasure? In answering these questions, Kant connects his transcendental anthropology of *feeling* with a critique of the power of *judgment*. The normative universality of aesthetic feeling is explicable in terms of purposiveness as an a priori principle of judgment that governs both experiences of beauty and our investigation of nature. By showing the connection between aesthetic feelings and purposiveness as a principle of judgment, Kant also solves the problems of incompleteness in the first and second *Critiques*.

Before showing how purposiveness explains the normativity of aesthetic judgment, Kant explains how it functions in the formation of empirical cognitions and thereby deals with an incompleteness in the transcendental anthropology of cognition offered in the *Critique of Pure Reason*. In section one, we saw how Kant argues for metaphysical principles such as universal causation, but in the *Critique of Judgment*, Kant points out that the a priori principles that make empirical cognition possible as such do not ensure the further "unity of experience . . . as a system in accordance with empirical laws" (5:183) because they do not preclude the possibility that each different experience of succession is governed by a different causal law.

[F]or nature in general (as the object of a possible experience) that law [of causation as such] is cognized as absolutely necessary . . .
[H]owever, the objects of empirical cognition are still . . ., as far as we can judge a priori, determinable in so many ways . . . that specifically distinct natures . . . can be causes in infinitely different many ways . . . Thus we must think of there being in nature, with regard to its merely empirical laws, a possibility of infinitely manifold empirical

laws . . . ; and with regard to them we judge the unity of experience (as a system in accordance with empirical laws) as contingent. (5:183)

The first *Critique* showed that changes in the world must happen according to causal laws, but it failed to show that the set of causal laws governing the world is finite, much less that these laws fit into anything like a systematic whole within which diverse particular laws are explicable in terms of more general laws (but cf. A642-68/B670-96). The first *Critique*, in other words, gave no indication of the possibility of anything even approximating the grand unified theories that scientists seek.

But since such a unity must still necessarily be presupposed and assumed . . . , the power of judgment must thus assume it as an a priori principle for its own use that what is contingent for human insight in the particular (empirical) laws of nature nevertheless contains a lawful unity. (5:183)

Human cognition involves seeing the natural world in terms of objects that change over time in accordance with necessitating natural laws, but we also seek systematic interconnections amongst these objects and the laws that govern them. While it would be consistent with the conditions of possibility of experience in general for each change to be governed by its own causal law, such that the world as a whole was a “manifold” (i.e., crazy mess) of different laws, human beings cannot actually think that this is the case. Our principles for investigating the world assume uniformity that, strictly speaking, we are not justified in assuming. Kant refers to “pronouncements of metaphysical wisdom” that are “scattered about in the course of science” such as that “Nature takes the shortest path” or “the great multiplicity of its empirical laws is nevertheless unity under a few principles” (5:182, cf. 5:185). Without such cognitive rules of thumb, we could never get anywhere in terms of a systematic empirical science; we would be left with the abstract metaphysical foundations of science laid out in Kant’s earlier transcendental philosophy.

For Kant, the “power of judgment” provides the transcendental basis for these scientific rules of thumb. In general, “the power of judgment is the faculty of thinking of the particular as contained under the universal” (5:179), and it has two basic forms: *determining* and *reflecting*. Determining judgment is the power to subsume a particular under an already-given universal. When one sees a particular tree as a tree, one *determines* that the particular present object falls under one’s concept of a tree. Reflecting judgment is the power by which we are able to “ascend from the particular . . . to the universal” (5:180). When one

reflects on the bones of an animal that is somewhat like a human being but not quite the same, one can gradually form the concept of a Neanderthal or Australopithecus. And whereas determining judgment has for its “rule” the (empirical) concepts that it applies,³⁹ reflecting judgment “requires a principle that it cannot borrow from experience, precisely because it is supposed to ground the unity of all empirical principles under equally empirical but higher principles” (5:180). The rules of thumb used in science are common-sense versions of this a priori (transcendental) principle of reflective judgment.

This transcendental principle of reflective judgment is purposiveness. “Nature specifies its universal laws in accordance with the principle of purposiveness for our faculty of cognition” in that natural laws are suited “for human understanding in its necessary business of finding the universal for the particular that is offered to it by perception and then further connection in the unity of the principle for all that is different” (5:186). That is, the otherwise “happy accident” that nature is suited to be understood as a systematic whole is required, a priori, as a “purposiveness in relation to the cognitive faculty of the subject” (5:185) that is assumed in every act of the regulative power of judgment. We treat nature as something *to be understood*, something that is *suited* to our cognitive faculties. (We do not have to assume that there literally *is* a designer who gave this purpose to nature. Nature is “purposive” – conducive to being understood – without literally having this “purpose.”)

For Kant, the assumed purposiveness of nature for our cognition is connected to our faculty of feeling pleasure (or displeasure). The connection is, at first, fairly straightforward: “the attainment of every end is combined with the feeling of pleasure,”⁴⁰ so if reflective judgment gives an a priori aim valid for everyone, “then the feeling of pleasure is also determined through a ground that is a priori and valid for everyone” (5:187). Specifically, since understanding particulars in terms of general laws and “bringing heterogeneous laws of nature under higher . . . laws” are demands of reflective judgment made possible through an assumed “purposiveness of nature for our understanding,” “if we succeed in this accord of such laws . . . , pleasure will be felt” (5:187-8). When the paleontologist studying a strange fossil is finally able to classify that

³⁹ Kant does not hold that there is a rule for the application of a concept. His point is that judgment is a specific power distinct from the power of understanding rules and distinct from the power (reason) of deriving one rule from another. The power of judgment is the power to apply rules to particulars, and this power to apply is not identical to the rule that is applied. But there must still always be some rule that is applied. In the case of determining judgment, this rule is given by the content of the concept.

⁴⁰ Importantly, Kant’s point here is not that one experiences pleasure only when one attains an end (see Allison 2001: 56, cf. Guyer 1979: 71), but that whenever one attains an end, one experiences pleasure.

fossil as a distinct species falling under some more general genus, she experiences pleasure at this success. The a priori principle of reflective judgment that makes possible the search for systematicity in our understanding of nature thus provides the first guide to a transcendental anthropology of feeling, since it proposes a necessary *end* for all human beings – unifying particulars under increasingly general laws – the attainment of which is a necessary and universal basis of pleasure for human beings. The presumption of purposiveness in nature grounds a necessary pleasure in actually discovering such purposiveness.

This pleasure is not aesthetic because it is both interested and conceptual. The pleasure is interested because it is a pleasure in accomplishing a specific goal. The goal is necessary and given by the structure of our cognitive faculties rather than by particular inclinations, but the pleasure is still an *interested* pleasure. The reason that the pleasure is conceptual is a bit more complicated. Insofar as one experiences pleasure in the purposiveness of the world through the systematization of one's experience, one's pleasure is pleasure *in empirical concepts*. And in these cases, the central dilemma of aesthetic judgment does not arise. The scientist who wants to explain why the fossil brings her pleasure can explain that this particular, which did not seem to fit into a systematic whole, can be understood according to such-and-such an empirical concept (Australopithecus, say, or Euleptes gallica). And even if one does not fully share in her pleasure – pleasure of discovery depends at least in part on working for it oneself – one can appreciate why pleasure is called for. In that sense, one can explain one's pleasure by reference to (empirical) concepts, and it is not purely aesthetic.

In principle, objects in the world might be purposive only in that they possess a general conduciveness to be understood. But in fact, Kant suggests the possibility of a purely “subjective” or “aesthetic” purposiveness, “the purposiveness of a thing . . . represented in perception . . . that precedes the cognition of an object, which is immediately connected with it *even without wanting to use the representation of it for a cognition*” (5:189, emphasis added). With the exception of the final clause, this description could apply to any reflective judgment, wherein one must *first* see an object (or a representation of it) as purposive and only *then* subsume it under general concepts. In this case, however, the purposiveness of the object is recognized without either subsuming the object under concepts or aiming for such subsumption. “The object is called purposive in this case *only* because its representation is *immediately* connected the feeling of pleasure” (5:189, emphasis added). Such a representation would be “an aesthetic

representation of purposiveness” (5:189). So far, however, this extension of the notion of purposiveness seems both arbitrary and mysterious. It is arbitrary because, as Kant immediately notes, it is still an open “question . . . whether there is such a representation of purposiveness at all” (5:189). And it is mysterious because is it not clear what such a representation would be purposive *towards* nor how any representation could be *immediately* connected with pleasure in this way.

At this point, Kant’s dilemma of aesthetic judgment has an important role to play. Just as Kant justifies the a priori categories of understanding as conditions of possibility of experience, and the postulate of freedom as a conditions of possibility of moral obligation, so here he uses the dilemma of aesthetic to show the nature and necessity of aesthetic representations of purposiveness.⁴¹ Along with his contemporaries, Kant takes for granted that there *are* normative judgments of feeling (“good taste”). Kant’s transcendental anthropology looks for conditions of possibility of such judgments. They are problematic, he argues, because they must be both subjective and universal. But having described the role of purposiveness in reflecting judgment that aims for systematic, empirical knowledge, Kant presents an account of what an immediately-felt, non-conceptual representation of an object’s purposiveness would have to be. And it turns out that such representations are precisely what would make possible normative claims that are both universal and subjective.

In particular, for Kant, an immediately-felt, non-conceptual representation of an object’s purposiveness would have to be a recognition of the suitability of an object not to any particular concept or concepts, but simply to humans’ cognitive faculties *in general*.

If pleasure is connected with the mere apprehension of the form of an object . . . without relation of this to a concept for determinate cognition, then . . . the pleasure can express nothing but [the form’s] suitability to the cognitive faculties that are in play in the reflecting power of judgment, insofar as they are in play, and thus merely a subjective formal purposiveness of the object. (5:189-90, cf. 5:222, 217)

The claim that, in the form of the beautiful, one’s cognitive powers are “in play” is central to Kant’s account. The “play” of these cognitive powers can be contrasted both with the *work* that such powers do (when, for example, reflecting judgment develops empirical concepts or unifies

⁴¹ See Ameriks 1982c, which defends Kant’s argument as a regressive one, like the argument of the first *Critique*.

diverse laws under more general ones) and with a possible *conflict* between such powers (such as when one's perceptions *resist* being brought under general concepts).⁴² For the feeling of beauty, the relevant cognitive powers are the imagination and understanding; when these powers play freely together, one feels aesthetic pleasure. What precisely is this "free play"? While there is substantial disagreement amongst commentators,⁴³ the general idea can be gleaned from Kant's examples of beautiful objects:

Flowers are free natural beauties. Hardly anyone other than the botanist knows what sort of thing a flower is supposed to be; and even the botanist, who recognizes in it the reproductive organ of the plant, pays no attention to this natural end if he judges the flower by means of taste. Thus the judgment is not grounded on any kind of perfection, any internal purposiveness to which the composition of the manifold is related. Many birds (the parrot, the hummingbird, the bird of paradise) and a host of marine crustaceans are beauties in themselves, which are not attached to a determinate object in accordance with concepts . . . but are free and please for themselves. Thus designs *à la grecque*, foliage for borders or on wallpaper, etc., signify nothing by themselves: they do not represent anything, no object under a determinate concept, and are free beauties. One can also count . . . musical fantasias (without a theme), indeed all music without a text. (5:229).

All of these examples refer to objects that inspire continuous reflection without getting to any determinate knowledge. Unlike clearly conceptualizable forms – such as an equilateral triangle (see 5:241) – that give no room to the imagination to examine them in new ways, a flower stimulates a constant redirection of attention from one aspect of its form to another, a constant attempt to reassemble the visually presented material with different emphases. But in contrast to a merely chaotic mish-mash of stimuli, the diverse perspectives that one can take on a flower are all orderly; the understanding is given constant encouragement to find patterns and generalities in the representations of the object. Moreover, the activities of imagination and understanding do not merely take place side-by-side; they are "reciprocally expeditious" (20:224). Finding patterns in one way of looking at a flower facilitates the re-presentation of the flower in yet another way, which leads to the recognition of a new order, and so on. One can continuously contemplate

⁴² See Allison 2001: 49-50. We will see another important case of conflict between powers when we turn to the sublime, below.

⁴³ For some discussions, see Allison 2001, Ginsborg 1997, Guyer 1979, and Zuckert 2007.

beautiful flowers, birds,⁴⁴ and musical improvisations, constantly reinterpreting them in the light of new “imaginative” ways of pulling together one’s impressions.

The purposiveness of beautiful objects is not towards *goals* of our cognitive powers (increasing knowledge of the empirical world) but towards the *activity* of those powers. For Kant, pleasure is a feeling of “the agreement of an object with the productive power[s] of the soul” (29:894); the “animation of [the] cognitive powers” of imagination and understanding gives rise to a pleasure, which “is itself” the consciousness of the purposiveness of the beautiful object (5:222). Beautiful objects are *pleasurable*, and because this pleasure lies in the *mere* animation of one’s cognitive powers, and not any *end* brought about by those powers, it is *disinterested*. This animation of cognitive powers is the effect of the mere representation of the object, not dependent upon any determinate cognition of the object, so one’s judgment that the object is beautiful is *non-conceptual* and thus *subjective*. But – and this is Kant’s key move – because the subjective basis of one’s judgment is the free play of cognitive powers *that all human beings share*, one can legitimately expect that *any* human being *should* feel pleasure at the representation of the beautiful object. Because the judgment that an object is beautiful is a judgment that the object is *purposive for one’s cognitive powers*, and because human beings share those cognitive powers in common, an aesthetic judgment carries with it a commitment to universality. (Of course, one might still get aesthetic judgments *wrong*. One’s pleasure in an object might only *seem* to be due to disinterested, non-conceptual contemplation. In that case, one might mistakenly call beautiful what is really agreeable or good.)

The purposiveness that grounds the subjective universality of aesthetic judgments of beauty also provides the basis for truly *free* pleasure. “among all . . . kinds of satisfaction only that of the taste for the beautiful is a disinterested and *free* satisfaction; for no interest, neither that of the senses nor that of reason, extorts approval” (5:210, cf. 5:354). As is typical for Kant’s transcendental anthropology, *freedom* in the context of pleasure is normatively governed. Just as free cognition is governed by a priori categories and forms of intuition, and free volition is governed by a categorical imperative, so the free experience of pleasure is governed by a principle of purposiveness by virtue of which one judges objects to be beautiful and hence worthy of pleasure when they properly enliven one’s imagination and understanding. Moreover, the free pleasure in beauty is a particularly *human* sort of pleasure:

⁴⁴ Crustaceans, too. See <http://www.reefkeeping.com/issues/2007-05/as/index.php>, accessed June 4, 2009.

“Agreeableness is also valid for nonrational animals; beauty is valid only for human beings . . . ; the good is valid for every rational being in general” (5:210).

Although aesthetic pleasure is free even from moral considerations, it provides an important “mediating concept” between morality and our sensuous nature (5:196, cf. 5:176). Aesthetic pleasure provides an “analogy” to morality (5:353) and “prepares us” for it (5:267). The universality, freedom, disinterested pleasure, and non-conceptual basis of aesthetic pleasure all correspond to similar traits of the moral law. And feeling pleasure in beautiful objects proves to human beings that we are not merely the playthings of our instincts and inclinations. Although aesthetic pleasure is not itself motivational (since it is disinterested), this possibility of disinterested pleasure reinforces the categorical demands of morality with a subjective (and pleasing) basis for thinking that one can be sufficiently free from inclination to meet those demands. Moreover, the subjective universality of taste promotes the development of cultures of taste, within which one seeks to conform one’s own aesthetic judgments to those of others. One’s effort to resist what Kant will later call “aesthetic egoism” is implicit in the universal nature of aesthetic judgment itself, and this development into a tasteful, social being prepares the way for overcoming the “moral egoism” that treats one’s own desires as more important than those of others.

In an important sense, then, Kant’s account of the feeling of pleasure in beautiful objects completes his transcendental anthropology. With this “critique of aesthetic judgment,” Kant presents the entirety of human mental life – cognition, volition, *and feeling* – as susceptible to transcendental investigation. Like cognition and volition, human feeling is normative and one can investigate the conditions of possibility of this normative structure from-within. While cognition is governed by a priori principles of the understanding and volition by an a priori principle of reason, feeling is governed by an a priori principle of *judgment*: the principle of purposiveness. Moreover, Kant uses his account of beauty to bridge the gap between nature and freedom in both the cognitive and volitional dimensions. With respect to cognition, the experience of beautiful objects involves reflectively judging about objects in the world and feeling the purposive suitability of this world to our cognitive capacities. This purposiveness of the world for reflective judgment provides a ground (albeit a subjective one) for regulating the investigation of nature in accordance with an assumption of its suitability for systematic understanding, thus bridging the gap between the *Critique of Pure Reason*’s assurance that the world would conform to certain general structures of human cognition and the need to be able to expand the

scope of knowledge systematically. With respect to volition, the experience of beautiful objects reveals, in the most subjective dimension of human existence, a universality and autonomy that is analogous to and preparatory for moral choice.

Kant *could* have ended his transcendental anthropology with his account of the beautiful, but he added two important dimensions to his *Critique of Judgment*: a theory of the sublime, and an account of teleological judgment. The account of the sublime is a natural addition to Kant's critique of aesthetic judgment. Especially with the publication of Edmund Burke's *A Philosophical Enquiry into the Origin of Our Ideas of the Sublime and Beautiful*, the issue of the relationship between aesthetic judgments of beauty and of the sublime became a hot topic in the 18th century. Like the experience of beauty, the experience of sublimity is something with which human beings find ourselves. When contemplating the Milky Way, or St. Peter's in Rome, or a mountain landscape, one can have a distinctive sort of experience of not being able to "take it all in." And while this inability to take it all in involves a certain kind of frustration, it can also be strangely satisfying. Similarly, when one experiences "[b]old, overhanging, and, as it were, threatening rocks, thunderclouds towering up in the heavens, . . . volcanoes with their all-destroying violence, . . . [or] the boundless ocean set into a rage," one can feel like one's "capacity to resist [is] an insignificant trifle in comparison with their power" (5:261). Again, however, this painful frustration of one's powers can be conjoined with a special sort of pleasure. Kant had "observed" and discussed these "sublime" feelings in detail almost thirty years before earlier,⁴⁵ but not until the *Critique of Judgment* does he try to give an account from-within of how such a feeling could be warranted.

In the *Critique of Judgment*, Kant transforms his early, casual account into a detailed theory of the relationship between these two forms of aesthetic judgment. The beautiful, for Kant, is the unambiguously pleasurable feeling of the harmony of our imagination and *understanding*. The sublime, by contrast, is the deeply ambivalent feeling of *disharmony* between our imagination and our *reason*. There are important similarities between the two:

The beautiful coincides with the sublime in that both please for themselves. And further in that both presuppose neither a judgment of sense nor a logically determining judgment but a judgment of

⁴⁵ In *Observations on the Feeling of the Beautiful and the Sublime* (1762).

reflection: consequently the satisfaction does not depend on a sensation, like that in the agreeable, nor on a determinate concept, like the satisfaction in the good . . . Hence both sorts of judgments are also *singular*, and yet . . . universally valid in regard to every subject, although they lay claim merely to the feeling of pleasure. (5:244).

The sublime, however, is an odd sort of universal pleasure because “that which . . . excites in us the feeling of the sublime may . . . appear in its form to be *contrapurposive* for our power of judgment, *unsuitable* for our faculty of presentation, and as it were *doing violence* to our imagination, and yet is judged all the more sublime for that” (5:245, emphasis added). Judgments of the sublime involve recognizing *disharmony* between the objects and our powers. In cases of what Kant calls the “mathematically sublime” (5:248f.), where we can’t “take it all in,” the disharmony is between human reason, which seeks “the comprehension of every appearance . . . into the intuition of a whole” (5:257), and the imagination, which is “inadequate” for this task (5:252). In cases of what Kant calls the “dynamically sublime” (5:260f.), the disharmony is between our natural capacities as a whole and a powerful, threatening object or situation. But if the beautiful pleases through the harmony it makes evident amongst our cognitive powers, how can the sublime, which makes evident a disharmony amongst those same powers, also please?

With the sublime, an initially painful disharmony gives rise to a feeling of pleasure when one recognizes in oneself a capacity that outstrips nature itself. With mathematically sublime objects, one sees that the demands of one’s (theoretical) reason outstrip nature, so that one seeks a unity in nature that cannot be satisfied by anything one’s senses can provide. And with the dynamically sublime, one’s recognition of the frailty of one’s natural life can reveal that one’s natural life does not exhaust who one is, that human nature includes a moral demand to have a good will that no natural forces can undermine. In both cases, the experience of certain natural objects gives rise to a feeling of one’s own transcendence over nature. Thus, Kant claims, “true sublimity must be sought only in the mind of the one who judges, not in the object in nature . . . That is sublime which even to be able to think of demonstrates a faculty of mind that surpasses every measure of the senses” (5:256, 250). In reflecting on certain objects, one comes to recognize a disharmony caused by the superiority of one’s humanity over the sensible, natural world. While this disharmony initially provokes displeasure, the *source* of the disharmony – one’s transcendent reason –

inspires an ambivalent, but nonetheless intense and pleasurable, feeling of self-esteem.

The feeling of the sublime completes Kant's account of aesthetic pleasure and supplements his treatment of the beautiful in three important ways. First, because Kant aims to give a complete transcendental anthropology, he must account for all of the ways in which humans' feelings of pleasure can be governed by a priori norms. Since humans' experience of the sublime is governed by such norms, it must be discussed. Second, the sublime provides an important balance to the contribution of the beautiful to Kant's anthropology of cognition, corresponding to the role that the transcendental dialectic plays in Kant's *Critique of Pure Reason*. There Kant showed that while the understanding structures our experience of the world, reason imposes demands that transcend any possible experience. In the beautiful, we feel the conduciveness of the world to human understanding; in the experience of the sublime, we feel how reason imposes demands that transcend the world. Finally, and perhaps most importantly, the sublime provides a bridge between nature and freedom that is importantly different from that of the beautiful. With the beautiful, humans' experience of fitness between themselves and nature makes us aware of a free, disinterested, universal capacity for pleasure that is analogous to moral demands. With the sublime, especially the dynamically sublime, humans directly feel their moral dignity. The experience of the sublime involves feeling precisely the same sort of respect for oneself that is constitutive of moral motivation. Thus while the sublime reflects the disconnect between oneself and nature, it also marks a bridge from an experience of nature that is not itself moral to a respect for oneself that plays a central role in moral motivation.

With his account of the sublime, there is an important sense in which Kant *has* completed his transcendental anthropology of feeling. The rest of the *Critique of Judgment* does not directly address humans' faculties of feeling at all. But for Kant, while "[i]t is . . . in taste . . . [that] the power of judgment reveals itself as a faculty that has its own special principle and thereby makes a well-founded claim to a place in the general critique of the higher faculties of cognition" (20:244), this a priori principle can be applied more widely than *merely* to govern aesthetic feeling:

once the capacity of the power of judgment to institute a priori principles for itself is granted, then it is also necessary to determine

the scope of this capacity, and for this completeness in critique it is required that its aesthetic faculty be recognized as contained in one faculty together with the teleological and as resting on the same principle, for the teleological judgment about things in nature also belongs, just as much as the aesthetic, to the reflecting power of judgment. (20:244)

Already, we have discussed one important non-aesthetic dimension of the power of judgment. Nature's purposiveness to our cognitive powers is not limited *merely* to free play but also enables an increasingly systematic *determinate* understanding of nature. But in the "critique of teleological judgment" (5:357), Kant goes further. Once "teleological judging" is "drawn into our research into nature" (5:360), one can consider the possibility of "objective purposiveness," wherein certain objects in nature are understood as natural purposes. Importantly, nature might not have included *objective purposes* (20:216) The laws of physics and chemistry are merely mechanical, though they show that matter is purposive for our cognitive powers because it can be subsumed under general laws. But Kant argues that we can think of a different sort of purposiveness of nature, what he calls "objective purposiveness":

Experience leads our power of judgment to the concept of an objective . . . purposiveness, i.e., to the concept of an end of nature, only if there is a relation of the cause to the effect to be judged which we can understand as lawful only insofar as we find ourselves capable of subsuming the idea of the effect under the causality of its cause as the underlying condition of possibility of the former. (5:366-7)

For objective purposiveness, the "natural laws" under which we subsume a given phenomenon depend upon thinking of causes of that phenomenon as *for the purpose* of their effect. If one can make sense of an organized being only as "a thing . . . [that] is cause and effect of itself" (5:370), then this being will be a "natural end." When one understands the motion of the heart in terms of its functional role in promoting the circulation of blood, and the circulation of blood in terms of promoting the life functions of an animal, and these life functions as in turn ensuring the continual motion of the heart, one interprets an animal teleologically, in terms of purposiveness. When, further, one sees an individual animal as both the effect of its species and the cause of the continuation of the species, one interprets the animal teleologically; it exists *for* the propagation of the species (and vice versa).

In theory, there might not be "natural ends," but in fact one finds self-propagating organized beings in the world "which cannot be

explained through [mechanism] alone” (5:374). The result is that human beings are entitled, and even required, to posit a principle for judging organized (biological) beings: “*An organized product of nature is that in which everything is an end and reciprocally a means as well.*” Or, in less technical lingo, “Nothing is in vain, purposeless, or to be ascribed to a blind mechanism of nature” (5:376). Importantly, for Kant, these principles are purely “regulative,” mere heuristics “for guiding research into objects of this kind” (5:376). But as heuristics, they are

indispensably necessary . . . In fact, [anatomists of plants and animals] could just as little dispense with this teleological principle as they could do without the universal physical principle, since, just as in the case of the abandonment of the latter there would remain no experience at all [as shown in the first *Critique*], so in the case of the abandonment of the former principle there would remain no guideline for the observation of a kind of natural thing that we have conceived of teleologically as an end. (5:376)

For Kant’s transcendental anthropology, the addition of these teleological principles has two important implications. First, it allows a limit to the causal explanation that the first *Critique* justified. While Kant insists that *in principle* everything in nature is explicable in terms of efficient causes and even that we are required to explain nature mechanically – that is, in terms of basic properties of matter – as much as possible (5:379, 429), he concedes that for humans studying the living world, such explanations will often not be possible. Second, Kant shows here a willingness to introduce new principles for judgment on the basis of empirical discoveries. The principle of objective purposiveness *precedes* and *guides* empirical research; biologists *assume* purposiveness prior to finding the specific purposes of particular aspects of organized beings. But this assumed purposiveness is itself the result of discovering *through experience* that certain beings in nature can only be understood (by us) in this way.

The need to investigate organized (living) things in accordance with a principle of purposiveness also gives rise to two further implications that will prove important for Kant’s anthropology as a whole. The first of these will be discussed in the next chapter. Briefly, just as Kant’s *Critique of Pure Reason* created a space for an empirical anthropology that views human beings as empirical objects subject to natural laws, his *Critique of Judgment* makes clear that, like other living things humans are irreducible to strict mechanism, requiring instead teleological explanation. In particular, Kant allows for basic biological powers that ground causal laws but the origin of which cannot be traced to any more

basic powers (say, of matter). The second implication is discussed in detail in the *Critique of Judgment* and constitutes the most important contribution of that work to the third key question of philosophy: “What may I hope?”

After explaining that organized beings in general must be understood as natural ends, Kant explains,

It is therefore only matter insofar as it is organized that necessarily carries with it the concept of itself as a natural end, since its specific form is at the same time a product of nature. However, this concept necessarily leads to the idea of the whole of nature as a system in accordance with the rule of ends . . . (at least in order to test natural appearance by this idea). (5:378-9)

Once one finds the fruitfulness of studying living beings in accordance with the principle that “Nothing is in vain,” it is natural to extend this heuristic principle to studying nature as a whole. This immediately yields fruitfulness in a scientific ecology studying organisms’ interdependence; one sees, for example, how plants nourish animals that in turn provide “for the human being, for the diverse uses which his understanding teaches him to make of all these creatures,” but also how, in turn, “plant-eating animals exist in order to moderate the excessive growth of the plant kingdom” and humans and other meat-eaters exist to keep the plant-eaters in check (5:426-7). But Kant also insists that this study of nature naturally leads one to think about what could be the “final end” of nature as whole. “A final end,” for Kant, “is that end which needs no other as the [teleological] condition of its possibility” (5:434), that is, something that we can see as being a self-sufficient *end-in-itself*. One can understand parts of organisms as teleologically-ordered towards the whole, and even individual organisms as ordered towards the species, but one can still ask “why do these creatures exist?” (5:426). Answering this ecologically only pushes the question back further, “but why do *those* creatures exist?” Naturally, Kant argues, we want an answer that justifies the whole world.

Given Kant’s transcendental anthropology of volition, the answer is both obvious and problematic. The answer is obvious in that Kant has already shown that there *is* something that is an end-in-itself: *humanity*. But the humanity that is an end-in-itself requires the transcendental freedom that grounds the possibility of a good will. And *that* cannot be an end *of nature* because neither transcendental freedom nor the good will are objects in nature; Kant establishes freedom and morality from a practical rather than empirical standpoint, as “things-in-themselves”

rather than “appearances.” Thus Kant insists that despite seeking a final end of nature, “if we go through the whole of nature, we do not find in it, as nature, any being that can claim the privilege of being the final end of creation” (5:426).

Fortunately, however, once we know that human beings as transcendently free choosers are the *final* end of nature, we can look for an “*ultimate*” end of nature that would identify “that which nature is capable of doing in order to prepare [the human being] for what he must himself do in order to be a final end” (5:430). Identifying this “ultimate end of nature” provides the basis for a rational hope that nature will cooperate with our moral vocation. In the *Critique of Judgment*, Kant focuses on the regulative principles that can guide our empirical study of nature. Given that humans’ *final* end is moral, Kant suggests that the *ultimate* end of nature, as that which nature can do to support our moral freedom, must be the human being’s “aptitude for setting himself ends at all and . . . using nature as a means appropriate to the maxims of his free ends in general” (5:431). Thus nature as a whole tends towards humans’ cultivation in “skill” at using things for their purposes, “discipline” whereby they rise above inclinations, civil societies that establish objectively right relationships amongst people, and even the emergence of cultures of taste, within which “beautiful arts and sciences” flourish (5:432-3). The *details* of Kant’s account of human beings as the ultimate end of nature, including the empirical evidence that emerge from (and in turn support) his regulative principles, emerge elsewhere and will be discussed in chapter three. But Kant’s *Critique of Judgment* shows how purposiveness as the principle of regulative judgment not only grounds aesthetic judgments but even leads, through its application to biology and ecology, to a conception of human beings as ultimate ends of a purposively ordered nature.

Kant’s *Critique of Judgment* is a transcendental anthropology of the faculty of feeling and the power of judgment that provides that faculty with its regulative principle. As an analysis of feeling from-within, the *Critique* shows how there can be non-conceptual normative standards for judgments of taste, and it reveals an analogy to morality in the most sensuous aspect of human nature: our feelings of pleasure. In its further analysis of purposiveness in the study of nature, this *Critique* not only justifies the assumption of order in nature but even shows how the assumption of purposiveness plays a necessary role in regulating humans’ study of objects with a certain kind of complexity. Throughout this *Critique*, Kant also emphasizes that purposiveness as a principle of

judgment is merely *regulative*, and in that sense it is more intimately human than either the cognitive principles of the first *Critique*, which constitute the structure of the empirical world, or the moral principles of the second, which apply to all rational agents. In the end, though, these intimately human principles provide a foundation for answering the question “What may I hope?” both affectively and rationally. In aesthetic pleasure, we legitimately feel hopeful in our cognitive strivings for systematic understanding of the world and in our moral aspirations for disinterested, universally-justifiable choices. And in our understanding of nature as a teleologically-ordered whole, we look for (and find) evidence that nature as a whole cooperates with our highest moral vocation.

V. Conclusion

Kant’s three *Critiques* present a picture of human beings as finite but free knowers, actors, and feelers. Human knowledge is constituted by passively received intuitions that are conceptualized by an understanding that spontaneously (that is, freely) imposes categories to cognize objects. Human action involves subordinating subjective and therefore finite maxims to an autonomous moral principle. And aesthetic pleasure arises from the free play of faculties that testify to our finitude. Moreover, the transcendental anthropology of volition in particular provides a (practical) proof that one is a *transcendentally* free “homo noumenon,” capable of acting on grounds that are undetermined by empirical causes. The transcendental anthropology of cognition ensures that the empirical *expression* of one’s transcendentally free choices will always be a “homo phenomenon,” susceptible to empirical description in terms of natural laws (6:417-8). And the transcendental anthropology of feeling shows how the empirically given world supplies material that provokes pleasurable aesthetic feelings that, in different ways, reveal our freedom to us.

Insofar as humans are homo phenomena, they must be understood in terms of categories of the understanding and forms of intuition. But even as homo phenomena, humans are still distinct from merely physical nature in that we are teleologically-ordered biological organisms with particular features, many of which have important implications for applying the moral law in practical life. Insofar as human beings are free homo noumena, we are both negatively free, in that our (noumenal) choices are not determined by any particular empirical causes, and positively free, in that we are subject to the moral

law as the law of our own will (autonomous). By virtue of our freedom, we are worthy of respect and hence the proper “end” of moral choice, and we are worthy of “awe” and hence proper objects of sublime feeling.

Most fundamentally, human’s noumenal freedom manifests itself in the freedom of choice that underlies moral responsibility, but all three of Kant’s critiques deal with human beings as free in the sense that humans are subject to normativity not only in action but also in thinking and feeling. All three realms of human life involve laws, norms, and requirements that are not *causal* but nonetheless perceived as binding from-within. Kant’s critical works thus not only set up the general framework of phenomenal-noumenal humanity but also specifically address the *noncausal* laws that govern human beings. Unlike the observation-based anthropology we’ll discuss in the next chapter, this transcendental anthropology is thoroughly normative throughout, a normative account of the human being from-within: an epistemology, an ethics, and an aesthetics.

However, “transcendental anthropology” is incomplete as an overall answer to the question “What is the human being?” The *Critique of Pure Reason* shows that empirical knowledge is possible and that reason pursues more and more completeness of knowledge of the empirical world. Because human beings appear in the empirical world, transcendental anthropology must be supplemented with an empirical anthropology that describes what humans look like “from-without.” Moreover, Kant’s *a priori* moral philosophy requires supplementation by an “empirical part” that will involve “judgment sharpened by experience” to know how the moral law should be applied and how “to provide [it] with access to the human will” (4: 388-9). And finally, while the practical postulates of God and immortality and the general teleology revealed through natural beauty and human biology give *some* basis for moral hope, Kant suggests that “experience and history” can provide further reasons that “we should not despair about our species’ progress toward the better” (7:329). It should thus come as no surprise that while Kant was developing his *transcendental* anthropology, he was also engaged in detailed *empirical* studies of human beings. Such empirical study is necessary to complete his answer to his question “What is the human being?” and thereby to fully answer his remaining questions: “What can I know [including empirically about human beings]?”, “What ought I do [to human beings with the empirical features that we have]?” and “What may I hope [based on the progress human being have made historically so far]?” It is to this *empirical* anthropology, then, that we now turn.

CHAPTER 2: KANT'S EMPIRICAL ANTHROPOLOGY

In the last chapter, we examined Kant's "transcendental" anthropology, his examination of the cognitive, volitional, and affective dimensions of the human being from the standpoint of a priori, normative, autonomously-given laws governing those faculties. But Kant also engaged intensely in empirical debates about human beings. The next three chapters focus on different dimensions of Kant's empirical anthropology. First, in this chapter, I examine Kant's overall empirical anthropology of the human mind, that is, his empirical psychology. This psychology includes Kant's accounts of the different faculties of human beings, the causal laws that describe the activity of those faculties, and the bases of such faculties in "natural predispositions" found in humans' biological nature. In chapter three, I turn to two more specific aspects of Kant's empirical anthropology, his treatments of human evil and of the historical nature of the human species. And in chapter four, I turn to Kant's accounts of human diversity.

I. The Possibility of Empirical Psychology

Given the importance of Kant's transcendental anthropology, he might seem merely to dismiss empirical anthropology. But in fact, one of the key claims that Kant establishes in his transcendental account of cognition is that human beings are capable of having empirical knowledge of their world, and he emphasizes that such empirical knowledge *includes* empirical knowledge of *human beings* themselves. His *Critique of Pure Reason* paves the way for the empirical study of human beings. Although Kant insists throughout the *Critique* that a human being cannot cognize itself "in accordance with what it is in itself," he constantly adds "that through inner sense we intuit ourselves . . . as we are internally affected by ourselves . . . [and thus] we cognize our own subject . . . as an appearance" (B156, cf. B69, 153). Like everything else we can cognize, human beings can be cognized as appearances, as "homo phenomenon" (MM 6:418). Even where the *Critique of Pure Reason* most emphasizes the possibility of human freedom, Kant insists that human beings insofar as they appear in the world are subject to empirical study: "all actions of a human being are determined in accord with the order of nature . . . [I]f we could investigate all the appearances . . . there would be no human action we could not predict with certainty" (A549/B577, cf. 20:196). Kant gives a striking example to illustrate this general point.

Let us take . . . a malicious lie First, we endeavor to discover the motives to which it has been due, and secondly, we proceed to determine how far the action . . . can be imputed to the offender. As regards the first question, we trace the empirical character of the action to its sources, finding these in defective education, bad company, in part also in the viciousness of a natural disposition insensitive to shame *We proceed in this enquiry just as we should in ascertaining for a given natural effect the series of its determining causes.* But although we believe the action is thus determined, we nonetheless blame the agent. (A554-55/B 582-83, emphasis added; cf. too 29:1019-20)

In the *Groundwork*, too, Kant reiterates that “everything which takes place [is] determined without exception in accordance with laws of nature” (4:455), and in the *Critique of Practical Reason*, he goes so far as to say that if we knew the relevant preconditions, “we could calculate a human being’s conduct for the future with as much certainty as a lunar or solar eclipse” (5: 99)

Despite Kant’s insistence on empirical study of human beings, there are three reasons that many call into question whether Kant can really allow for empirical anthropology. First, it seems impossible for Kant to admit that humans are susceptible to truly “scientific” study, since any such study must provide *universal* claims about its objects (as, for example, Newton’s laws do about matter), which Kant seems to rule out for empirical anthropology. In his *Metaphysical Foundations of Natural Science*, Kant even claims, “The empirical doctrine of the soul can . . . never become . . . a science of the soul, nor even a psychological experimental doctrine” (4:471).⁴⁶ Second, a completely empirical-causal anthropology might seem to conflict with Kant’s very strong claims about human freedom (see chapter one).⁴⁷ Third, empirical anthropology just seems fraught with practical problems. Kant discusses epistemic challenges such as the fact that “if a human being notices that someone is . . . trying to study him, he . . . either . . . cannot show himself as he really is or . . . does not want to be known as he is” (7:121) and moral dangers of self-study, such as that “self-observation . . . is the most direct path to illuminism or even terrorism, by way of a confusion in the mind of supposed higher inspirations and powers flowing into us . . .

⁴⁶ E.g. Gouax 1972, Westphal 2004.

⁴⁷ Cf. Reath 1989, Baron 1995.

from who knows where” (7:133).⁴⁸ These comments suggest that even if empirical human science is possible in principle, it is unreliable and dangerous in practice.⁴⁹

In fact, however, none of these concerns preclude an empirical anthropology suitably construed. To start with the first point, Kant rejects the possibility of an empirical human science in a very strict sense of “science” (4:471, 20:238, 28:679). Kant uses the term “science” to refer only to knowledge that is a priori. Newton’s laws count as science, for Kant, because “outside of what lies in [the empirical] concept [of matter], no other empirical principle is used” (4:470).⁵⁰ By contrast, even if one starts with empirical, psychological concepts – such as the existence of a mind – one cannot derive further substantive claims about the mind a priori.⁵¹ And unless one can derive claims about human psychology a priori, one cannot make strictly universal claims: “Experience tells us, indeed, what is, but not that it must necessarily be so, and not otherwise. It therefore gives us no true universality” (A1-2, cf. B3-4, A91/B124, A196/B241). General claims about human beings, if based merely on experience, might seem to be “merely fictitious” (A196/B241). In fact, however, Kant’s rejection of a scientific and strictly universal status for empirical anthropology does not preclude the possibility of what he calls a “historical systematic natural doctrine of the inner sense” (4:471), a “natural science . . . improperly so called, . . . [which] would treat its object . . . according to laws of experience” (4:468). Kant even compares certain psychological forces to “motive force[s] in the physical world” (25:577), and he remarks that because “physics is knowledge of the object of outer sense, and the knowledge of human beings as the object of inner sense, . . . [empirical anthropology] deserves . . . to be treated as a science in academia, just as is physics” (25:472-3). Kant does not here mean that empirical anthropology will be a science in the same sense as physics, since it will not have a mathematical, a priori foundation. The claims of empirical anthropology will, correspondingly, not be necessary claims. But even in the first Critique, Kant admits a kind of universality that experience can provide: “empirical rules . . . can acquire through induction . . . comparative

⁴⁸ The idea here is that “we make supposed discoveries of what we ourselves have carried into ourselves” (7:133), acting as though the apparently involuntary flow of our thoughts is due to some sort of supernatural inspiration.

⁴⁹ Cf. Wood 1999, Loudon 2000.

⁵⁰ As we will see in chapter one, Kant holds that Newton’s laws can be derived a priori from the concept of matter and the a priori structures of human cognition.

⁵¹ Kant gives several reasons for this, including the fact that because the form of inner sense is time, “which has only one dimension,” “mathematics in inapplicable to the phenomena of inner sense” (4:471) and the fact that “observation itself already changes and displaces the state of the observed object” (4:471).

universality, that is, extensive applicability” (B124/A91, see too B 3-4). While not “science” in Kant’s strict sense, empirically-grounded laws of human beings constitute a comparatively-universal, systematic presentation of human mental and social life.

The second concern about the possibility of Kantian empirical science arises from Kant’s account of human freedom. Many commentators have rightly pointed out that the account of free action in Kant’s transcendental anthropology of volition offers a different conception of moral psychology than the traditional empiricist belief-desire model according to which human beings are simply motivated by their strongest active desire. But those who take this insightful alternative approach to conflict with Kant’s empirical account of human action⁵² are misguided. As we will see in chapter one, Kant’s transcendental idealism aims to show how an empirical and even causal model of human behavior leaves room for real freedom, not within the empirical model but as something distinct from (and grounding) humans’ empirically-knowable character. Kant’s transcendental anthropology of cognition shows that humans experience everything – including themselves – in terms of a structure of cognition that interprets change in terms of causal relationships. But this metaphysics of the empirical world leaves room for a different standpoint from which freedom is possible, and Kant’s transcendental anthropology of desire – his moral philosophy – makes clear that human agents must see themselves as free causes of their actions. This implies, of course, that Kant’s empirical anthropology is only empirical; it does not provide access to what Kant will call the human being as it is in itself, the “homo noumenon” (MM 6:418, cf. 7:397-400). It is possible to have access to what the human being is like in itself (as we will see in chapter one), but empirical anthropology provides no such access.

The final challenge for rigorous empirical investigation of human beings is the set of specific difficulties with self-study that make empirical anthropology – as Kant put it at the opening of his first course in anthropology – a “hard descent into the Hell of self-knowledge” (25:7). As he says in his published *Anthropology*,

[A]ll attempts to arrive at such a [human] science with thoroughness encounter considerable difficulties that are inherent in human nature itself.

⁵² One such commentator has said, “if the moral law determines choice by exerting a force that is stronger than the alternatives, moral conduct will result from the balance of whatever psychological forces are acting on the will . . . It is not clear that this model leaves room for any real notion of will or choice” (Reath 1989: 290-91). Cf. too Baron 1995:189, Westphal 357-8, and even Wood (2003: 50).

1. If a human being notices that someone is observing him and trying to study him, he will either appear embarrassed (self-conscious) and cannot show himself as he really is; or he dissembles, and does not want to be known as he is.

2. Even if he only wants to study himself, he will reach a critical point, particularly as concerns his condition in affect, which normally does not allow dissimulation: that is to say, when the incentives are active, he does not observe himself, and when he does observe himself, the incentives are at rest.

3. Circumstances of place and time, when they are constant, produce habits which, as is said, are second nature, and make it difficult for the human being to judge how to consider himself, but even more difficult to judge how he should form an idea of others with whom he is in contact; for the variation of conditions in which the human being is placed by his fate or, if he is an adventurer, places himself, make it very difficult for anthropology to rise to the rank of a formal science. (7:120-1)

And in a draft of his anthropology, Kant adds,

[T]he I which has been observed by itself is a sum total of so many objects of inner perception that psychology has plenty to do in tracing everything that lies hidden in it. And psychology may not ever hope to complete this task and answer satisfactorily the question: “What is the human being?” (7:398-99)

For Kant, empirical study of human beings proceeds by means of both introspection and observation of others, and both of these forms of study face several of the problems Kant describes here: mental life is intrinsically complex, human beings typically act differently when being observed, self-observation is inhibited by the fact that many of the most interesting and important activities in human life preclude the calm and attentive work of introspection,⁵³ and human beings can develop contingent characteristics – habits of time and place – that seem essential. We could add even more difficulties today, such as unconscious motivation or the fundamental attribution error.⁵⁴ The

⁵³ When one is overcome with anger, for example, it is impossible to actually observe the nature of that anger because anger prevents the calm interest and introspective attention involved in self-study. And when the anger has passed, one can no longer see – and may not accurately remember – what the anger looks like in action.

⁵⁴ See chapter eight for more.

result of all of this can be a desperation about the possibility of ever (empirically) answering the question, “What is the human being?”

Despite these cautions, Kant regrets that while “[n]othing seems more interesting . . . than this science, . . . nothing is more neglected” (25:7) and insists that “an anthropology . . . that is systematically designed” is possible and “yields an advantage for the reading public,” including the promotion of “the growth of [this] science for the common good” (7:121-2). Kant even maintains that anthropology begins with “general knowledge of human beings” (7:120) and “is provided with a content by inner sense” (7: 398, cf. 25:252, 863-5). And while Kant warns about the dangers of such introspection, he insists that it can be done in a way that is relatively free from danger, claims that observing oneself is a duty (6:441-2), and gives both general advice and specific examples of how to introspect well. Moreover, Kant insists that this introspection provides only a basis for further study; one must make use of and accordingly adapt general knowledge through interpreting others.

If we want to judge about other people, we must alter our point of view, namely

1. transpose my point of view and then
2. put myself in the other’s point of view . . . To take a point of view is a skill which one can acquire by practice. (25:475)

Reading good literature – Kant suggests Shakespeare, Montaigne, and Fielding⁵⁵ – further cultivates and supplements proper self-observation. Travel and the reading of travel literature provides further material for reflection on both human nature in general and “local knowledge” of the varieties of human beings (see 7:120). In the end, disciplined self-study supplemented by careful study of others and reading literature by those skilled in putting themselves into others’ points of view can alleviate the problems faced by any attempt “to observe human beings and their conduct, [and] to bring their phenomena under rules, [which] is the purpose of anthropology” (25:472). While Kant doubts the possibility of a wholly satisfactory empirical anthropology, he aims to develop as full an empirical account as possible, or at least a sufficient account “from which a prudent use in life can . . . be drawn” (25:472, cf. 7: 119).

⁵⁵ In other contexts, Kant includes such authors as Samuel Richardson, Moliere, the English Spectator (by Allison and Steele), and Rousseau. For references to literature as a source of anthropology, see 7: 221, 25:473. For a fascinating recent study of Kant’s readings of Milton, see Budick 2010.

Kantian empirical anthropology, then, is general rather than strictly universal, and thus a science only in a loose sense. Even as such a science, it is particularly vulnerable to error given humans' tendencies to get ourselves wrong. But a fallible quasi-science laying out empirically-justified general laws of human beings is, for Kant, possible, interesting, and useful. The rest of this chapter lays out the overall framework of this empirical anthropology. The next two chapters get into more specific topics in Kant's empirical anthropology: human evil, history, and diversity.

II. Kant's Faculty Psychology

When Kant began working on empirical anthropology, the dominant approach to the empirical study of human beings in Germany (promoted by Christian Wolff and Alexander Baumgarten) involved laying out different human mental states in terms of various "faculties of soul" and then showing how these faculties could be reduced to a single faculty of "representation." The idea was that rational cognitions were the clearest and most distinct representations of the world, and sensory cognitions, feelings, and desires were representations with varying degrees of obscurity and indistinctness. Against this view, many philosophers and emerging psychologists (most prominently, Moses Mendelssohn, Johannes Tetens, and August Crusius⁵⁶) argued that human mental states were irreducibly distinct; the main alternatives suggested were either a very wide diversity of human mental states, a bi-partite model within which belief and desire are irreducible to each other, or some combination of these within which irreducibly distinct mental states can be classified broadly into beliefs and desires. Although he lectured from a textbook (Baumgarten's) that promoted the single-faculty approach, Kant's own work defended mental state pluralism. Unlike his compatriots, however, Kant insisted upon combining a broad mental state pluralism with a fundamentally tri-partite structure. The basic idea was that there are a large number of irreducibly distinct sorts of mental states that can be grouped into three irreducibly distinct types: cognition, feeling, and desire.

Kant's argumentative strategy for this view is two-fold. First, he develops a general philosophy of science according to which one should seek to "deriv[e] diverse powers, which we know only through observations, as much as possible from basic powers" (28:564, cf. 8:180-1; 28:210; 29:773-822; A648-9/B676-7). One should assume as many basic powers as are really necessary, a point Kant emphasizes by

⁵⁶ Cf. Crusius 1745, §§73, 444; Watkins 2005:91; Hatfield 1990.

comparing Descartes, who “explains all [physical] phenomena from the shape and the general motive power of bodies,” with Newton’s “more satisfactory” method that allows the assumption of “certain basic powers . . . from which the phenomena are derived” (29:935-6, cf. A649-50/B677-8). The phenomena one finds in both the physical and mental worlds require more than a single basic power. So while Kant seeks to reduce powers as much as possible (for instance, by showing that memory is a form of imagination), his focus is on not overly reducing mental powers. Second, Kant lays out some specific arguments to show that particular mental powers are irreducible to one another. For example, Kant emphasizes his isolation of feeling as a state distinct from both cognition and desire/volition, noting that feeling is not merely a confused cognition of a thing, and emphasizing that aesthetic pleasures, no matter how intense, do not give rise to volitions (29:877-8). He points out that while cognition is “related merely to the object and the unity of the consciousness of it,” a volition is “the cause of the reality of this object” (20:206). More generally, Kant argues that one can only reduce distinct powers if one can find a power from which they “could be derived” (8:181n). In the end, Kant insists, “there must be several [basic powers] because we cannot reduce everything to one” (29:773-822).

For Kant, the set of distinct basic powers includes each of the five senses; an “inner sense” by virtue of which we are aware of our own mental states; the imagination; higher cognitive powers of reason, understanding, and judgment; a power of feeling pleasure and displeasure; and various powers of volition. Kant groups these distinct powers into the general faculties of cognition, feeling, and desire, and further sub-divides them between “higher” and “lower” faculties. “Lower” faculties are primarily receptive, while “higher” faculties are “self-active” (28:228, 29:880, 28:584), by which Kant does not mean the transcendental freedom of the homo noumenon but a “comparative concept of freedom” according to which “actions are caused from within” (5:96).⁵⁷ We can lay out Kant’s overall taxonomy of mental powers as follows:

Faculties of soul

	Cognition	Feeling	Desire/Volition
--	-----------	---------	-----------------

⁵⁷ For a discussion of the relation between these sorts of freedom, see Frierson 2008.

Higher	Judgment, Understanding, Reason	“satisfactions or dissatisfactions which depend on the manner in which we cognize the objects through concepts”	“Motives” based on practical principles
Lower	Sight, hearing, taste, smell, feeling, inner sense, imagination (including memory)	“satisfactions and dissatisfactions which depend on the manner in which we are [sensibly] affected by objects” (28:254)	“Stimuli,” impulses rooted in instinct or inclination.

2. Causal Laws Governing Human Beings

Kant did not discuss the structure of human mental faculties simply to argue against Wolff’s reduction of the mind to a single faculty of representation. Getting clear on different mental faculties is crucial for developing a full empirical anthropology because “the concept of cause lies in the concept of power” (28:564) and in empirical human science we seek “natural laws of the thinking self” based on “observations about the play of our thoughts” (A347/B405, cf. 25:472). Each distinct mental power is governed by its own causal laws (including laws governing how it relates to other mental powers), and a complete empirical anthropology describes these laws.

For Kant, faculties of soul are causally ordered such that “pleasure precedes the faculty of desire, and the cognitive faculty precedes pleasure” (29:877-8). Moreover, “all desires have a relation to activity and are the causality thereof” (25:1514, cf. 29:1024); desire plays the same role in psychology that motive forces like momentum play in physics (25:577). In fact, desire is defined as a representation that is the ground of an action that brings about some state of affairs (6:211, 399; 7:251; 29:1012), so there are no actions not preceded by and caused by desires, and no desires that do not lead to actions (in the absence of external impediments). For any human action, a sequence of causes can be traced as follows:

Cognition □ Feeling of pleasure (or pain) □ Desire □ Action

At any step along this progression, the causal chain could be cut off. For example, when a normal human being tastes a mango (cognition), that taste gives pleasure (feeling), that pleasure causes a desire for the mango, and that desire leads one to eat (or continue eating) the mango. But one's mango might be snatched away, preventing one from eating (or continuing to eat) the mango. Or one might see a beautiful flower (cognition) and experience a "disinterested" pleasure that gives rise to no subsequent desire. Or one might learn that the capital of Iceland is Reykjavik and thus have a cognition, but without this cognition giving rise to any pleasure or desire.

This sketch requires filling in. Kant needs to explain what gives rise to cognitions in human beings, how and when those cognitions give rise to pleasures, and how and when those pleasures give rise to desires. As Kant offers the details of these causal laws, his account gets extremely detailed, so here I only highlight aspects of his account. With respect to the lower faculty of cognition, Kant distinguishes between the five traditional senses (sight, hearing, touch, taste, and smell), "inner sense" (our ability to "observe" our own inner mental states), and imagination (which, for Kant, includes memory, foresight, and the imagination as a "fictive faculty" to think of things that we never experience). About the five senses, Kant lays out only the most general descriptions, such as that "the sense of touch lies in the fingertips and the nerve endings (papillae) and enables us to discover the form of a solid body by means of contact with its surface" (7:155) and that sight, touch, and hearing are "mechanical" while taste and smell are "chemical" (25:495). Kant speculates about "a faculty of the nerves [that] underlies the mind . . . in separating and combining given sensory representations" and even suggests that some sort of "water of the brain" that is encountered by the "ends" of the "stimulated optic nerve" or "auditory nerve" (12:34). But Kant's dominant approach is basically Newtonian, that is, not to try to explain how light, for instance, causes a visual sensation by stimulating the optic nerve, but simply to classify what physical causes bring about this mental state. Just as Newton does not solve the "problem" of gravitational action at a distance but instead names and classifies the phenomenon, Kant does not try to "solve" the so-called "mind-body problem" of how states of the brain cause mental states, instead merely classifying the basic powers that underlie these connections.

The most extensive psychological discussion of lower cognitive faculties comes with the imagination. Kant posits that imagination is

governed by three fundamental laws: affinity, forming intuitions in space, and association. Association, for example, is the principle that “empirical ideas that have frequently followed one another produce a habit in the mind such that when one idea is produced, the other also comes into being” (7:176). Hearing a particular song may trigger thoughts of the person with whom one often listened to that song, or the thought of a certain book may cause one to think of the place where one read that book. The imagination also figures centrally in Kant’s account of language: it is by virtue of customary association between sounds and thoughts that those sounds (and eventually written words) come to stand as symbols for those thoughts.

Kant’s discussion of the higher faculty of cognition is more complex. This faculty is subdivided into three basic powers: reason, the understanding, and judgment. Most generically, “Understanding draws the general [i.e., concepts] from the particular Reason draws the particular from the general The power of judgment is the subsumption of one concept under others [or of particulars under concepts]” (29:890). The power of judgment operates according to the principles governing analogy – “things . . . which . . . agree in much, also agree in what remains” – and induction – “what belongs to many things of a genus belongs to the remaining ones too” (9:133, see too 24: 772). The understanding generates certain concepts as an immediate consequence of sensory perceptions, but most concepts of the understanding are generated through chains of comparison, reflection, and abstraction. With respect to the former, Kant argues that “on the occasion of experience” certain “concepts have arisen through the understanding, according to its nature” (28:233), such as the basic concepts of causation and substance that make it possible for our experience to be intelligible as experience of an objective world.⁵⁸ In other cases, sensory cognition leads to empirical concepts, such as when seeing “a spruce, a willow, and a linden” leads one to

compare these objects with one another [and] note that they are different from one another in regard to the trunk, the branches, the leaves, etc.; but next . . . reflect on that which they have in common among themselves, trunk, branches, and leaves themselves, and . . . abstract from the quantity, the figure, etc. of these; and thus . . .

⁵⁸ Even in the case of these basic concepts, one “understands” them only in an implicit sense, sufficient for applying them for cognizing the world. Coming to articulate one’s implicit understanding of causation in terms of the generic concept “cause” takes a process of “reflection” (28:233) on the experience that they help to structure. (Small children make use of causal inferences long before they understand what the word “cause” means.)

acquire a concept of a tree. (Jasche Logic, 9:95; see too 24: 252-3, 753, 907)

Reason, finally, operates through principles of logic: the cognition of the premises of an argument give rise to a cognition of the conclusion of that argument. When I think about the facts that “Socrates is a human” and “All humans are mortal,” I am led to the thought that “Socrates is mortal.”

Thus far, Kant’s account of how higher faculties work tracks how they ought to work, but Kant knows that people’s higher faculties often do not function according to these ideal laws, and he develops an account of how “other activities of the soul . . . are connected with the judgments of the understanding” to generate a “mixed effect” that can be mistaken for “a judgment of the understanding” (16:283-4).⁵⁹ Such mixed effects result from what Kant calls “prejudices,” which function as alternative principles by which some cognitions give rise to others according to causes distinct from the understanding strictly speaking. For example, “the prejudice of the prestige of the age” leads some to favor the writers of antiquity more than they should, thereby “elevating the relative worth of their writings to an absolute worth” (9:79). For those affected by this prejudice, cognitions of claims associated with a particular ancient writer will immediately give rise to affirmation of those claims, a transition inexplicable in terms of properly functioning higher cognitive powers alone. Prejudices primarily arise from “imitation, custom, and inclination” (9:76), and Kant especially highlights the role of intellectual laziness in cultivating prejudices. Prejudices do not wholly displace higher cognitive faculties, but they provide a way for Kant to make sense causally of transitions between beliefs that are not actually justified, and thus cannot be explained in terms of the higher cognitive faculties alone. This account of prejudice, supplemented with detailed accounts of various prejudices and an account of how the higher and lower faculties of cognition relate, make up his overall account of the faculty of cognition.

Kant’s account of the faculty of the feeling of pleasure and displeasure is the most original, complex, and confusing aspect of his faculty psychology. The originality lies in Kant’s claim the faculty of feeling can be reduced to neither cognition nor volition, and Kant’s basis for this claim is his account of aesthetic pleasure. Even those who

⁵⁹ My account of deviations from proper functioning of the higher cognitive faculties here focused on those deviations that take place in ordinary human knowers. Kant’s account of mental illness will be discussed in chapter four.

argued against Wolff's attempt to reduce all the basic powers of the soul to a single one generally ended up describing pleasure as either a subjective form of cognition like color or scent or as a constitutive part of desire. But Kant explains,

We have pleasure or displeasure without desiring or abhorring, e.g. if we see a beautiful area, then it enchants us, but we will not on that account wish at once to possess it. Pleasure or displeasure is thus something entirely different from the faculty of desire. (29:877)

The difference between pleasure and volition cannot be explained merely in terms of strength. Even a mild pleasure in the thought of a sweet treat brings with it a desire to eat (more of) that treat, while the most intense purely aesthetic pleasure gives rise to no volition at all. Because there can be pleasures that are not connected with volition at all, Kant argues, pleasure cannot be seen as merely a component part of volition. But pleasure is also a sort of feeling towards an object that cannot be reduced to cognition of it. Someone could understand everything there is to know about an object that pleases me and still not find pleasure in it. Pleasure indicates something about *me*, not necessarily anything about the object. So while many feelings might be intimately linked with cognitions and volitions, feeling in general cannot be reduced to these cognitive-volitional aspects. Thus Kant articulates what precisely feeling is in its own right.

On Kant's general account of feeling, there can be very different kinds of feelings, but all feelings are, in some sense, feelings of satisfaction (pleasure) or dissatisfaction (displeasure).⁶⁰ Because pleasure is not a kind of cognition, Kant rejects the dominant (at the time) Leibnizian-Wolffian definition of pleasure as "the [obscure] sensible representation of the perfection of an object" (20:226, cf. 5:227). Instead, Kant offers two "definitions" of pleasure:

- (1) "The consciousness of the causality of a representation with respect to the state of the subject for maintaining it in that state can here designate in general what is called pleasure" (5:220, cf. 20:230, 15:241; 25:459, 785; 28: 247, 586; 29:890; 6:212, 7:231)
- (2) "Pleasure is the representation of the agreement of an object or of an action with the subjective conditions of life" (5:9n; cf. 5:204; 7:231; 15:246, 15:252, 16:133, 25:167-8, 181, 1501; 28:247, 586, 29:891).

⁶⁰ Kant sometimes uses these terms synonymously, and sometimes distinguishes pleasure as sensuous satisfaction, where satisfaction is a more generic term. (So God can feel satisfaction but not pleasure, and we find satisfaction but not pleasure in obeying the moral law.)

First, pleasure is defined simply as a mental state (a “representation”) oriented towards preserving itself. The feeling of pleasure just is a self-persistent mental state. But Kant relates this feeling oriented to persisting in one’s state with the concept of “life,” which he connects with self-activity and sometimes defines as a “faculty of a being to act in accordance with laws of the faculty of desire” (5:9n, cf. 28:275, 680). In the case of human beings, “life” involves the full set of mental powers of cognition, feeling, and desire. The general idea is that when something seems to promote the activity of one’s powers, a distinctive mental state arises that reflects this advancement of activity; this mental state is called “pleasure.” When something seems to inhibit activity, one feels displeasure. In a lecture on metaphysics, Kant connects his two definitions:

The feeling of the promotion of life is pleasure, and the feeling of the hindrance of life is displeasure. Pleasure is when a representation contains a ground for being determined, for producing again the same representation, or for continuing it when it is there. (28:586)

Thus we might say that when one feels pleasure, one feels like continuing in one’s state because one’s state seems conducive to the activity of one’s powers. When one feels displeasure one feels like ending one’s state because one’s state feels like an inhibition of activity.

Given his definition(s) of pleasure, Kant divides possible objects of pleasure into different categories. Most fundamentally, and central to Kant’s insistence that pleasure is not merely an aspect of desire, Kant claims that pleasures can be distinguished into those that give rise to desires and those that do not. The key to this distinction is that desires are “objective” in that they are directed towards bringing about their objects in the world, while pleasures, in themselves, are wholly subjective, both in that they reflect something about the subject (whether one’s overall state is conducive to life) and in that they aim to preserve themselves subjectively (as mental states). Generally, preserving pleasurable states involves acquiring objects or objective states of affairs that bring pleasure. Pleasure in a mango depends upon actually eating the mango. Such pleasures, in order to “produce again the same representation or . . . continue it” (28:586), give rise to desires, mental states that actually affect the world by causing one to act (e.g. eat the mango). Kant calls such pleasures “interested” or “practical” and proposes that other pleasures – aesthetic ones – are not interested: “[A] judgment about beauty in which there is mixed the least interest is very partial and not a pure judgment of taste” (5:205). Kant takes this point quite far, claiming that if “the palace that I see before me [is] beautiful” I

will feel a distinct pleasure in the contemplation of it, even if “were I to find myself on an uninhabited island . . . and could conjure up such a magnificent structure through my mere wish, I would not even take the trouble of doing so” (5:204-5, cf. 29:878). Aesthetic pleasures arise from reflection, and the actuality of the object of one’s reflection is not necessary in order to promote the “free play of the powers of representation” (5:217) that grounds the feeling of pleasure. Kant’s aesthetics (discussed in chapter one) focuses on a transcendental account of these disinterested pleasures. For the purpose of his empirical account, his main purpose is to show that there are such pleasures and thereby distinguish the faculty of feeling from that of desire.

With the exception of aesthetic feelings, pleasures sustain themselves by means of the faculty of desire. The faculty of desire is the most complicated faculty in terms of its causal laws, and explaining it fully requires a discussion of the biological roots of causal laws in Kant’s account. In general, though, the faculty of desire is “the faculty to be, by means of one’s representations, the cause of the objects of those representations” (5:9n, 6:211) or “the self-determination of a subject’s power through the representation of something in the future as an effect of this representation” (7:251). The key point here is that desire is a mental state by virtue of which one becomes a cause of the objects of that mental state. Whereas cognition merely thinks about its objects and feeling merely enjoys its objects, desire actually brings about what it represents.⁶¹ To desire something is to have the requisite mental state for bringing that thing about. Even when desire is not fully self-conscious (as in animals, or as with bare urges) or when it is merely a response to sensory stimuli, it still represents a mental state directed towards an object as a cause of bringing that object about. When desires are more deliberate and self-conscious, when they follow from the higher faculty of cognition such that we want something because we understand what it is, then Kant describes such desire as “a faculty to do or to refrain from doing as one pleases” (6:213). When one has a desire, one might also lack the ability to actually bring about the end of that desire. A desire, for Kant, always involves a volitional commitment to an object, but when one is committed to bringing about the object while still recognizing that one lacks the power to actualize that commitment, one’s desire is called a mere “wish” (6:213). When one desires an object and is also aware of one’s power to bring about that object, one “chooses” it (6:213). It should be clear, here, that “desire” is in some ways closer to what we consider “choice” than it is to what we typically consider “desire.” When a person

⁶¹ Thus, contra Wolff, the differences between these faculties, as well as differences between higher and lower faculties, are not reducible to differences in clarity and distinctness.

“desires” something in Kant’s sense, it means that they have the sort of mental state that will bring about its object if it can. Thus what we might consider a mere “desire” would for Kant be an inactive *ground* for a possible desire. “Concupiscence (lusting after something) must be distinguished from desire itself, as a stimulus to determining desire. Concupiscence is always a sensible modification of the mind but one that has not yet become an act of the faculty of desire” (6:213).

Beyond this general description of desire, Kant must explain why it is that desires arise for certain objects and not others. Generally speaking, the account is fairly simple. Given a cognition that gives rise to a practical pleasure, one will experience a desire for the object of that pleasure. The problem is explaining why certain cognitions give rise to practical pleasures while others do not. Kant’s solution to this problem is, justifiably, extremely complex. Generally speaking, he distinguishes between higher and lower faculties of desire based on whether they are moved by pleasures in higher cognitions (principles informed by concepts or ideas) or lower cognitions (brute sensations or imagination). But within these classifications, Kant must explain the variety of forms of human desire, and he articulates that account in the context of a description of biological and environmental factors that characterize human beings. To get a fuller empirical anthropology, then, we need to turn to Kantian human biology.⁶²

3. Human Natural Predispositions

One of the central claims of Kant’s biology is his claim that “it would be absurd . . . to hope that there may yet arise a Newton who could make comprehensible even the generation of a blade of grass according to natural laws that no intention had ordered” (5:400). In rejecting a “Newton of a blade of grass,” Kant denies one dominant strand of 17th and 18th century biology, theories of mechanistic “epigenesis” that aimed to explain the origin and nature of life in terms of purely physical forces.⁶³ But Kant also rejects the dominant alternative, “preformationism,” which assumed that all humans (and other animals) pre-existed in the egg or sperm of their most distant ancestors, formed in miniature and waiting to emerge. The dominant concept in Kant’s biology is the “natural predisposition” (*Naturanlage*), a concept that combines important aspects of both epigenesis and preformationism. Natural

⁶² Ultimately, even Kant’s biology is insufficient for understanding human nature, since human beings are essentially an historical species, so one must turn to his account of human history; see chapter three.

⁶³ Proponents included Descartes, Hobbes, and LaMettrie.

predispositions are “grounds of a determinate unfolding which are lying in the nature of an organic body” (2:434). Kant argues that “chance or universal mechanical laws could not produce such agreements [adaptive homologies], [so] we must consider such arrangements as preformed . . . [and] the mere faculty to propagate its character is already proof that a particular . . . natural predisposition for it was already to be found in the organic creature” (2:435). But Kant still insists not only that “outer things can well be occasioning causes” (2:435) for the development of these predispositions but also that “even in the case of the structure of an animal, it can be assumed that there is a single predisposition that has the fruitful adaptiveness to produce many different advantageous consequences” (2:126). Like epigenesists, Kant wants to explain natural variety using the smallest number of explanatory principles, but like preformationists, he allows that some elements of biological structure cannot be explained by mechanism alone.⁶⁴ Moreover, the way in which Kant suggests that outer things affect the development of natural predispositions ends up being selective rather than purely mechanical. That is, natural predispositions “lie ready . . . to be on occasion either unfolded or restrained, so that [an organism] would become suited to his place in the world” (2:435). Kant’s use of predispositions is thus a sort of adaptationist⁶⁵ account of biological (including human) development: organisms are born with a set number of predispositions that develop in response to various environmental conditions in accordance with what is needed to thrive within those conditions.

By appealing to “predispositions” in living beings, Kant does not commit himself to any particular metaphysical conception of the development of living things. Instead, he aims to effect an epistemic and methodological shift. Unlike epigenesists, who try to account for the emergence of biological structures from simpler processes, Kant argues that investigation of living beings proceeds best when one seeks to discover the minimal number of predispositions from which one can best explain the full range of biological phenomena one finds in the world. One corollary of this epistemic use of predispositions is that Kant’s pessimism about a Newton of a blade of grass is not a denial that the generation of a blade of grass may in fact be causally determined according to mechanistic laws, but only an admonition to distinguish biology from physics and allow forces in the former that might be inadmissible in the latter (see 5:411, 415, 422).

⁶⁴ As we will see in chapter eight, there is a sense in which current biology does account for natural predispositions in terms of purely physical causes. What for Kant are inexplicable natural predispositions turn out to be genes that evolved through processes of natural selection.

⁶⁵ One might even say “proto-Darwinian” here. See chapter eight.

This gives rise to a further aspect of Kant's biology. Given that organic predispositions serve purposes within organisms, Kant adds, as a "heuristic principle for researching the particular [biological] laws of nature" (5:411), a "principle of final causes" (5:387) "in order to supplement the inadequacy of [mechanical explanation] in the empirical search for particular laws of nature" (5:383). According to this heuristic principle, "nothing in [an organized product of nature] is in vain, purposeless, or to be ascribed to blind mechanism" (5:376). By relegating teleology to the status of a heuristic, a "maxim of the reflecting power of judgment" (5:398) that "is merely subjectively valid" (5:390), Kant can adopt a biology that explains natural organisms in terms of purposive structures. Predispositions that are not further explainable either physically (through mechanist epigenesis) or supernaturally (through divinely ordained preformationism) are susceptible to a teleological analysis. One can biologically explain why these developmental possibilities and not others are present through explaining what purpose they serve. The result is that in Kant's biology, one can legitimately ask, about any biological structure, what purpose that structure serves, and answers to such questions are legitimate parts of biological investigation.

Kant's preformationism has several important implications for his empirical anthropology. First, it allows Kant to forego describing how human predispositions came into existence: "we begin with something that human reason cannot derive from prior natural causes – that is, with the existence of human beings," including all of their natural predispositions (8:110). Kant's empirical anthropology reduces given powers to as few natural predispositions as possible, explaining environmental factors that allow certain predispositions (but not others) to flourish in (certain) human beings and using this small number of natural predispositions to explain what we observe of human beings. Second, Kant's emphasis on teleological explanation of these predispositions gives him additional resources for "explaining" predispositions without mechanistically explaining them. Kant gives teleological explanations of phenomena as diverse as sleep (7:166, 175, 190), laughter (7:261), distinctions between the sexes (7:305), and that "illusion" by which someone "who is naturally lazy" mistakes "objects of imagination as real ends" (7:175, 274). Third, Kant's preformationism contributes to his general disinterest in giving materialist explanations of psychological predispositions. Kant contrasts his approach to the empirical anthropology of his contemporary Ernst Platner, which Kant identified with "subtle, and . . . eternally futile inquiries as to the manner in which bodily organs are connected with thought" (10:145). Through positing predispositions as fundamental concepts in biology, Kant's

empirical anthropology can focus on explaining diverse mental phenomena in terms of as few basic powers as possible, tracing these basic powers back to purposive natural predispositions and the environmental influences that cause these predispositions to unfold, without being preoccupied with finding the physical structures that underlie those predispositions.

In laying out his empirical anthropology, then, Kant takes basic powers to be developed forms of natural predispositions. These natural predispositions provide bases for connections between mental states, grounds in human beings for observed laws covering such connections. Thus for any two mental states, we can describe their connection in terms of a causal law that is grounded in a basic power, which is itself the determinate unfolding of a natural predisposition.

Mental State 1 \square Mental State 2

↑

Actualized Natural Predisposition

The concept of natural predisposition allows Kant to expand the sense of “basic power” beyond the limited and abstract structure of his empirical psychology. Especially in the context of the faculty of desire, Kant develops a vocabulary for natural predispositions that provides the flexibility and variety needed to make sense of the myriad different ways in which human beings can be motivated to actions.

With respect to the faculty of cognition, Kant’s treatment of natural predispositions is fairly straightforward. The senses, inner sense, the imagination, and the higher faculties of judgment, understanding, and reason are all different natural predispositions in the human being (A66, 6:444-5, 25:1172, 29:915). Humans have natural predispositions to sense, imagine, and think in accordance with the laws described above. Thus in explaining the connection between one cognition and another, one appeals to the natural predispositions active in effecting that transition. For example, when the transition from the thought of one’s dog to the thought of dog food is effected by the imagination, one can describe this transition as follows:

Thought of dog \square thought of dog-food



Imagination (the predisposition governed by the law of association)

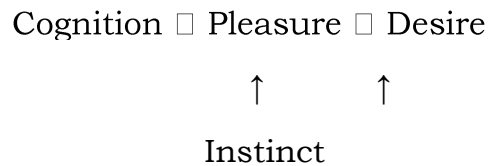
By contrast, the transition from the thought of one's dog to the thought "animal" would be effected by the understanding, hence the relevant predisposition would be different. In both cases, however, a complete explanation of the origin of a particular cognition must include, for Kant, not only the prior state that caused the cognition and the causal law according to which that state caused that transition, but also the natural predisposition that is the ground of that law.

There are variations amongst human beings in terms of the exercise of natural cognitive predispositions. Some of these are rooted in variations amongst predispositions themselves, such as certain forms of mental illness. Others involve a deficiency in the development of natural predispositions. And others, including all prejudices, involve circumstances in which some predispositions (linked either to imagination or to the faculty of desire) override the understanding and reason, leading to erroneous judgments. There are also positive variations in cognitive powers, such as wit or originality of thought, which Kant calls "talents," a sort of "excellence of the cognitive faculty which depends not on instruction but on the subject's natural predisposition" (7:220). Altogether, Kant develops an account of cognitive predispositions that identifies the basic powers of cognition as predispositions and then accounts for variations in cognitive abilities through either hereditary or acquired defects in these predispositions or their expression.

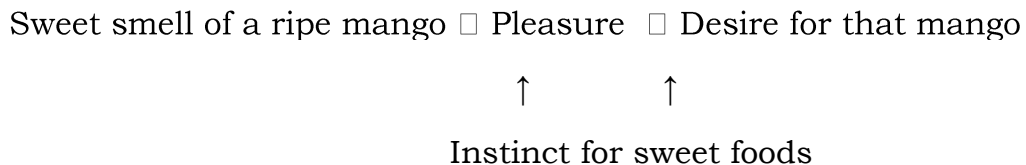
Predispositions become more important and complex with respect to the faculty of desire and the practical pleasures related to one's desires. As noted in the last section, the causal structure that determines whether particular cognitions give rise to desire or aversion can be exceedingly complex. Many things that give rise to desires in one person do not do so in others, things can give rise to desires sometimes and not others, and humans – even as objects of empirical study – seem capable of a kind of freedom of choice that might seem to preclude causal explanations. In every case, desires are preceded by cognitions that provoke feelings of pleasure that in turn provoke desires. But while virtually all feelings of pleasure cause desires for their objects (the only exceptions being the special cases of aesthetic pleasures), cognitions can

cause pleasure, displeasure, or no feelings at all. Kant seeks to trace this complex volitional structure back to two kinds of natural predispositions: instincts (*Instinkte*) and propensities (*Hänge*).

The nature and role of instincts is fairly straightforward. Among natural predispositions present in human beings are a set of instincts that ground connections between various cognitions and practical pleasures (or pains) that give rise to desires (or aversions) for objects of those cognitions. Given the distinctness between the faculties of feeling and desire, there would be, strictly speaking, separate predispositions underlying the connection between, on the one hand, a particular cognition and subsequent feeling, and, on the other hand, that feeling and its consequent desire. But because all practical pleasures give rise to desires and Kant offered an account of non-practical desires that explains how they cause feelings without subsequently generating desires, his detailed account of human motivation conflates the power that grounds a connection between cognition and feeling and the power that grounds the connection between the feeling and desire. Kant ascribes the transition from cognition to desire to a single basic natural predisposition. (For ease of presentation, I sometimes drop the reference to the intermediary practical feeling in Kant’s account and simply describe the role of natural predispositions as relating cognitions to desires.) In the case of instinct, Kant’s model of motivation maps straightforwardly onto his account of predispositions in general.



For example,



Often, instincts become operative when one is in the presence of the object that one’s instinct predisposes one to desire (or avoid): “little chicks already have from nature an instinct of aversion to the hawk, of which they are afraid as soon as they merely see something fly in the air” (28:255). With respect to human beings, Kant explains how smell, by

means of “its affinity with the organ of taste” and “the latter’s familiar sympathy with the instruments of digestion,” serves as an “instinct” that “guided the novice . . . allow[ing] him a few things for nourishment but for[bidding] him others” as though it were a kind of “faculty of pre-sensation . . . of the suitability or unsuitability of a food for gratification” (8:111). Central to these operations of instinct is that instincts ground connections between cognitions and anticipatory pleasures that give rise to desires. And these connections occur even before any experience of pleasures that might follow from the satisfaction of the desire. In cases where the objects of instincts are not present, Kant even suggests that instincts can be “directed to an indeterminate object; they make us acquainted with the object” (25:584). As he illustrates,

One knows that children, who are hardly born, show an instinct for nutrition, without knowing what they need, and immediately carry out the art of the physical law to suckle the breast; if they did not have the instinct, but one first had to accustom them to this, then many would perish. . . . We can see that the sexual instinct is a natural instinct by the fact that, even if they were in the monastery, when [the time of] puberty comes, persons are still disturbed by the instinct, and feel the need for an object which they do not yet know. (25:584)

One can have instincts with definite objects of present awareness, but one can also have instincts with indeterminate and unknown objects, instincts that agitate to activity in such a way that one comes into the presence of their objects. Moreover, the power of imagination can greatly expand the scope of instinct. Kant describes a scene where “a fruit which, because it looked similar to other available fruits which he had previously tasted, encouraged him to make the experiment” of eating it (8:111).⁶⁶ Given an association between a particular visual experience and a particular olfactory experience, a similar visual experience will – by virtue of the laws that govern the imagination – give rise to an imaginative idea that corresponds to that olfactory experience. Given a sufficiently strong instinctual connection between that olfactory experience and practical pleasure, the mere sight of a similar fruit will give rise to a desire to consume that fruit.

⁶⁶ In this essay, Kant emphasizes the cooperation between imagination and reason in the extension of one’s natural instincts and inclinations. However, one can also conceive of an extension that is purely due to imagination, as I suggest here.

Even with this expanded conception of instinct, however, most of human life is not directly governed by instinct, for two important reasons. First, much of what humans desire is not reducible to particular instincts. Human desires for the company of one's friends, wearing fashionable clothes, resting on comfortable sofas, watching one's favorite television programs, attending baseball games, and even for things like smoking cigarettes and eating fine foods, cannot be explained by appeal to brute instincts. These are all, in varying degrees, connected with habits that give rise to desires for certain objects. Second, even when we pursue objects for which we have instincts, humans typically do not pursue those objects directly from instinct. Instincts give rise to what we might call a desire, but we have a capacity to reflect on whether or not to pursue the object of that desire. Humans frequently decide not to follow through on instinct for the sake of something else, often something for which they do not have particularly strong instinctual desires at that moment. When I decide not to eat that delicious ice cream because I know that it will make me sick later, I do not act from any instinctual desire for long term health. If humans acted only from instinct, the task of explaining human motivation would require merely a catalog of relevant instincts and careful descriptions of environments in which those instincts play out. But human behavior is, as Kant recognized, much harder to explain.

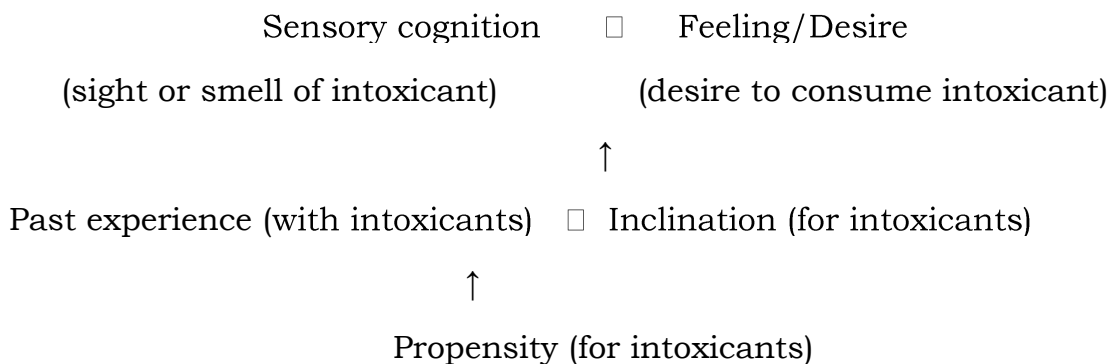
One might be tempted, at this point, to appeal to human freedom as a reason for the difficulty of explaining human behavior. And many have thought that the complexity of human motivation provides some support for Kant's account of freedom. But within his empirical anthropology, Kant takes the complexity of human action not as a reason to posit any kind of transcendental human freedom, but rather as a basis for a much more complicated but still wholly empirical anthropology. Kant adds the requisite complexity through a generous use of the category of a "propensity." In one lecture, Kant defines a propensity as a "natural predisposition" that provides "the inner possibility of an inclination" (25:1111-2; cf. 7:265, 25:1517). More generally, a propensity is a natural predisposition that does not itself provide a ground for connections between cognitions and practical pleasures (and thereby desires), but that makes it possible for the human being, in the context of environmental factors, to develop a ground for such connections. Having introduced this notion of a propensity, Kant puts it to use to address the two problems mentioned in the previous paragraph.

First, Kant focuses on human propensities for "inclinations," which he identifies as "habitual grounds of desire" (25:1114) and which, for the

purposes of his empirical anthropology, are distinguished from instincts.⁶⁷ Like instincts, inclinations provide bases for connections between cognitions and desires. Unlike instincts, however, inclinations are not natural predispositions but rather tendencies brought about through certain experiences. For example,

[S]avages have a propensity for intoxicants; for although many of them have no acquaintance at all with intoxication, and hence absolutely no desire for the things that produce it, let them try these things but once, and there is aroused in them an almost inextinguishable desire for them. (6:29; cf. 25:1112, 1339, 1518)

In some cases, one needs only a single experience of an object for an inclination to be awakened. Generally, however, inclinations require “frequent repetition” (25:1514). Kant adds that there is a generic propensity to develop habits, such that when one experiences something consistent over a long period of time, one develops an inclination for it (cf. 9:463-4). In any case of inclination, however, it is not enough to simply have exposure to something to develop an inclination for it. Experiences give rise only to inclinations when human beings already have requisite propensities. The model for explaining human action in those cases looks like:



In these cases, the immediate explanation for why a particular cognition gives rise to a practical pleasure and thereby a desire will be similar to the case of instinct, but because inclinations are not themselves innate, the account requires an extra level of complexity. And this complexity

⁶⁷ Within his moral philosophy, when Kant refers to “inclinations,” this term includes instincts as well. Kant adds, in his empirical anthropology, a special category of inclination called “passion.” I discuss passions in chapter six.

provides for much of the richness and diversity that one finds in human desires. Fancy clothes, comfortable sofas, cigarettes, and baseball are all possible objects of inclination, even when we have no instinctual desire for them. And because humans differ in their experiences, even those with the same propensities (and Kant allows for some, but not much, variation in basic human propensities⁶⁸) will end up with very different patterns of desire. A general propensity for competitive sport (or, even more generally, for esteem and physical exertion), leads to widely varying inclinations depending on the particular sports to which one is first exposed. Because propensities are natural predispositions, Kant does not give mechanical accounts for them, but he does aim to reduce the number of posited propensities to as few as possible; ideally, he would also provide teleological explanations for each propensity.

Kant also suggests that inclinations generally involve pleasure in ways that differ from instinct. For both instinct and inclination, experience of the object of desire brings a subsequent pleasure distinct from the practical pleasure that causes the desire. For instincts, this subsequent pleasure plays no explanatory role in the development of the instinct. The instincts for nursing or for sex motivate human beings to seek milk or sex innately, not because one has experienced their pleasures already. Instincts ground pre-sensations (8:111) of pleasure. But in the case of inclinations, the anticipatory practical pleasure that gives rise to desires generally follows from past experiences of the pleasure that one experiences when one attains the objects of desire. One accidentally experiences some object, gets pleasure from the experience, and forms an inclination that grounds future connections between the cognition of that object and the desire to experience it. One might taste an intoxicating beverage out of thirst or conformity (rather than a desire for intoxicants) or might literally fall into a pleasantly cool pool of water on a hot day. When the experience of such objects brings pleasure, one will seek intoxicating beverages even when one is not thirsty, or one will intentionally seek out and immerse oneself in cool pools of water. In these cases, we might specify the past experience as past experience of

⁶⁸ Even in the case of intoxicants, Kant's famous reference to the propensity of "savages" for intoxicants is more likely an assertion of the *universality* of this propensity than a limitation of it to a particular group. Kant's view seems best captured by his claim in a lecture on anthropology that "Human beings across the whole world have a propensity to drink [alcohol]" (25:1112). Elsewhere he uses "northern peoples" (25:1339), "the wildest peoples" (25:1112), "nations that have wine" (25:1518), and even (with respect to intoxicants more generally) the "people in Kamtschatka, [who] have a certain cabbage, which when they eat it, works in them a kind of madness, for which they love to have it" (25:1518). The point of these examples is not to pick out any particular group as uniquely susceptible to intoxicants, but to show that a propensity to drink that is undeniable in the case of Europeans is equally present in savage, or "raw" (*rohe*), people.

pleasure in the objects. Of course, one need not always experience pleasure in order to form an inclination. One who has started smoking can find herself craving cigarettes even while the actual experience of smoking is still generally unpleasant, and one who develops a habit of acting in a particular way can develop an inclination to continue acting in that way, even if it is not, in itself, particularly pleasurable. Generally, however, a propensity brings about a corresponding inclination at least in part through pleasure in attaining its object.

The addition of inclinations to Kant's account of human motivation greatly enriches that account, and it makes it possible to explain why there is such a wide range of divergent human interests. But inclinations, like instincts, still do not involve the reflective desires that characterize much human action. Kant captures this limitation by ascribing both instinct and inclination to the "lower" faculty of desire. Both affect human beings insofar as we are motivated by sensory or imaginative mental states, but not insofar as we govern our actions by means of concepts, principles, or maxims (the "higher" faculty of desire). For Kant, the higher faculty of desire, to which Kant assigns the term "choice" (6:213), "cannot be determined to action through any incentive except so far as the human being has incorporated it into his maxim" (6:24).⁶⁹ To explain how "maxims," or cognitive principles, can give rise to volitions and thereby actions, Kant cannot merely appeal to instincts or inclination. Instead, he appeals to yet another propensity, a propensity to what he calls "character."

Kant uses the term character in several senses throughout his writings. In the broadest sense, the character of a thing is the "law of its

⁶⁹ This passage is typically, and rightly, used to analyze the practical perspective from which we are transcendently free vis a vis all of our actions. Kant's specific language in this passage confirms that he primarily has this transcendently free power of choice in mind, rather than the specifically empirical faculty of choice. But the claim about acting on the basis of maxims here also has an empirical correlate. Empirically, the higher faculty of desire is the human capacity to have volitions that result from consideration of maxims, or principles of action, rather than mere stimuli. Insofar as one's volition falls under the higher faculty of desire, any lower incentives based on instinct or inclination must be conjoined with the recognition of a principle of action in order to motivate.

Seeing the connection between the transcendental importance of incorporating incentives into maxims and the empirical nature of higher volition highlights the distinctive role of maxims for Kant's account of motivation. What essentially distinguishes higher from lower volitions is that the motivating cognitions for lower desires are particular, while the motivating cognitions for higher desires are general. Thus, for example, Harry Frankfurt's discussion of first and second order desires (in Frankfurt 1988) does not track Kant's distinction between higher and lower cognitions, since one could – at least in principle – have general first order desires and particular second-order desires. For Kant, the *generality* of a desire – its lawlike form – is distinctive of human (rational) agents. For Frankfurt, it is second-order desires.

causality, without which it would not be a cause at all,” such that “every effective cause must have a character” (A539/B567, cf. 25:634). In this sense, gravity reflects the “character” of matter, and one’s instincts are part of the “character” of one’s lower faculty of desire. In a quite different sense, Kant uses “intelligible character” to refer to the free ground – “which is not itself appearance” – of one’s appearances in the world (A539/B567). Character in this sense has no role to play in empirical explanations of action, although Kant argues that this intelligible character grounds the empirical character of the higher faculty of desire. The character that plays an important role in Kant’s empirical theory of the higher faculty of desire is distinct from though grounded in intelligible character, and more specific than the character of an efficient cause in general. Kant defines this sense of character as “that property of the will by which the subject has tied himself to certain practical principles” (7:292) or “a certain subjective rule of the higher faculty of desire” (25:438, cf. 25:277), and this sense of character plays the same role for the higher faculty of desire that instincts and inclinations play for the lower. As Kant explains, “the man of principles, from whom we know for sure what to expect, not from his instinct, for example, but from his will, has character” (7:285, cf. 25:1514).

One can describe such motivations as follows:

Cognition (of a principle for action) □ Pleasure/Desire

↑

Character

“Character” is a matter of commitment to various principles or “maxims” of action. Thus, one may have a commitment to the principle “early to bed, early to rise.” In such a case, one’s actions might be explained as follows:

“Early to bed...” □ Desire to go to bed

↑

Fixed commitment to “Early to bed, early to rise”

Of course, this example is too simple in several respects. For one thing, the cognition of the principle “Early to bed, early to rise” is not in itself sufficient to generate the desire to go to bed, since one must also have awareness of the fact that it is evening – time for bed – rather than morning or afternoon. In order for one’s character to ensure that the principle will be efficacious in generating its corresponding action, one requires both perception of one’s situation – the sky is growing darker, the clock says 9 PM, etc – and the consciousness of the relevant principle.⁷⁰

Moreover, one’s commitment to the principle “early to bed, early to rise” is itself the result of other causes. Kant needs an account of the causes of character as such, that is, the ability to act in accordance with principles at all, and an account of the origins of the particular principles upon which individuals act. Regarding the first, Kant’s account of character development is similar to his account of the development of inclinations. There is a “propensity to character” (25:1172, cf. 25:651, 823, 1176) that is actualized by various experiences (cf. 25:1172; cf. 7:294), such that one might have the propensity but lack character, just as one might have a propensity to intoxicants but never develop the inclination. In the case of character, habit does not play a role in its formation. Instead, Kant emphasizes the role of education (25:1172), examples (7:294), and “moral discourses” (25:1173n1, cf. too 9:492-3), and he gives specific recommendations regarding the kinds of education that are most effective, such as avoiding “imitation” (25:635, cf. 7:325; 5:154; 25:599, 722ff., 1386) and using “discipline” (9:449) that can “first clear away the passions” so that character can develop without hindrance (9:486). Beyond these direct influences, Kant suggests oblique factors that play roles in character cultivation, such as stable and just political regimes, peace, and even progress in the arts and sciences. He even suggests that politeness cultivates character by combating passions and promoting self-control.⁷¹ And finally, Kant points out how other natural predispositions (especially temperament) facilitate character development (cf. 7:285, 290; 25:1388). All these elements work together

⁷⁰ These elements might themselves be causally related, and often are. When a practical principle comes to mind, one may be led to look for the presence of its conditions of application and thereby come to perceive those conditions. Alternatively, the perception of conditions of application (a darkening sky, a late clock) may remind one of one’s practical principles.

⁷¹ For more on the role of politeness in cultivating character, see Brender 1997, 1998, and Frierson 2005.

to transform a mere propensity into an active ability to govern oneself with conscious principles rather than reactive instincts and inclinations.

Many of the influences responsible for the development of character as such also foster specific practical principles, but Kant emphasizes that most of these principles still “rest on sensibility, and . . . merely the means for arriving at the end are presented by the understanding” (28:589). For example, one might learn “early to bed . . .” through instruction, but this instruction is effective because it proposes a plausible principle for satisfying one’s instincts and inclinations. Even in the absence of specific instruction, one with experience can form principles of action based on what actions best promote desired ends. Such principles are intellectual rather than sensible, but they still “rest on sensibility” because one formulates them for the sake of “lower” (i.e., sensible) inclinations and instincts. Even actions described as following “from inclination” are generally grounded in a character committed to principles that make objects of inclination its ends.⁷² Generally, inclinations for sweets do not directly cause one to eat them; rather, one understands that eating this food will satisfy a felt inclination, and (because of one’s character) this thought causes one to eat it.

Actions motivated by these “impure” principles of character are explained by an extremely complicated motivational picture. Through natural higher cognitive powers, sensory data are transformed into a conceptual understanding of one’s situation. At the same time, by virtue of one’s instincts and inclinations, one’s sensory awareness of one’s situation gives rise to various lower desires.⁷³ The understanding then provokes the thought of one or more practical principles based on how reason connects its conception of one’s situation with one’s felt lower desires. Thus one who recognizes the darkening sky under the concept of “early evening” might be led to think of the principle “early to bed...” by virtue of understanding this as the time at which going to bed will best facilitate the satisfaction of various inclinations over the long term. These practical principles give rise to practical pleasures and thereby desires –

⁷² Most human desires flow from principles to which we are committed in order to satisfy the instincts and inclinations of our lower faculty of desire. Sometimes these connections will be straightforward: human beings in wealthy nations typically consume food not immediately from instinct but from principles according to which we recognize the eating of food to be both immediately worthy of pursuit (because pleasurable) and ultimately useful for providing nourishment. Even foolish consumption of junk food is generally not directly instinctual but is a deliberate effort to satisfy the cravings of instinct according to principles – “Snickers really satisfies” – that we incorporate into the character of our higher faculty of desire.

⁷³ Or, more strictly speaking, proto-desires. If one reserves the term “desire” for the active desires that constitute real volitions (as Kant often does), then these are not yet desires in the strict sense.

which Kant, in these cases, calls “choices” – by means of a character that has been formed through education, social-cultural influences, one’s own past behavior, and the cooperating or hindering influence of inclinations and instincts. Both character in general and the inclinations that largely determine the content of the principles on which one acts are grounded in natural propensities. Thus human beings, due partly to different natural propensities but largely to different past experiences, will be motivated by similar sensory data to behave in different ways.

As complicated as this picture is, Kant thinks that human motivation is even more complicated, for three important reasons. First, the account given above assumes that for any given set of sensory data, there is only one way in which one’s natural powers can conceptualize that content and, more importantly, that this conceptualization only lends itself to a single practical principle. But it might well be that the recognition of the darkening sky is conjoined with a recollection of an invitation to a social gathering that promises to be particularly enjoyable. Here one may be led to think of the principle “early to bed, early to rise” but also the principle “don’t forgo opportunities for enjoyable social gatherings” (cf. 6:473, 7:277-82), when one cannot in fact act in ways that follow from both practical principles. In such cases, even one with a well-formed character will have conflicting possible grounds of action. One’s character could enable the former principle to give rise to a practical pleasure that would motivate one to stay home, or it could enable the latter principle to give rise to a practical pleasure that would motivate one to go out. From within practical reflection (transcendental anthropology), what one does is a matter of free choice. But empirical anthropology must provide a psychological explanation. Kant first insists, “in empirical psychology, wholly equal incentives cannot be thought” (28:678; cf. 25:278) because in the case of equal incentives, there would be no choice and thus no action (29:902). As a result, Kant distinguishes “living” and “dead” grounds of desire. Even when one has only a “dead” ground, one might still be left with something like a desire, with what Kant calls a “wish,” where the “ground determining one to action . . . is [not] joined with one’s consciousness of the ability to bring about its object” (6:213), in this case due to one’s pursuit of other ends. Thus one goes to bed because one’s overall character subordinates the principle of socialization to that of prudent rest, but one falls to sleep wishing that one could somehow both go to bed early and partake in the enjoyable party.

The second added complication to this account of choice is that although, strictly speaking, character requires commitment to act from consistent principles, very few people have character in this fully

developed sense. In a lecture, Kant specifically mentions difficulty with the practical maxim “early to bed, early to rise”:

[one] who is not steadfast in this, often lays hold of a resolve, of which he knows for sure that nothing will come, because he knows that he has already often broken resolutions. Then the human being is in his [own] eyes a wind-bag. He no longer has any confidence in himself . . . This is how it is with things for which one wants to break one’s habit . . . , such as sleeping in; for it is always said, just one more time, but then no more, and thus one again philosophizes oneself free of one’s plan . . . (25:624)

Sometimes inclinations directly overpower one’s higher faculty of desire, such that in the strict sense, one acts on the inclination alone, without the reflection that characterizes choice. But such cases are rare. More often, inclination corrupts the grounds of choice and one “philosophizes oneself free of one’s plan” by acting on a maxim that differs from what one had resolved. For Kant, this tendency is quite common. Truly firm character “is fixed very late,” only “com[ing] at a ripe old age” (25:654, 1385, cf. 7: 294). Most people have a kind of “bad” or “flawed” character (*schlechte Charackter*. 25:650, 1172; *Fehlerhafte im Charackter*. 25:1172). Such “character” is a “constitution of one’s higher powers” (25:227) according to which, rather than acting on from fixed principles, one allows the principles of choice to vary based on inclinations active at the time of choice. Here inclinations and instincts not only affect to which practical principles one commits oneself but also determine whether and to what extent those affect principles deliberation at particular moments. One with a firmly established character decides, by assessing the impact of various principles on her life as a whole, how to prioritize such principles. When the time comes for action, which practical principles determine action are set by this prioritization. One with flawed character might similarly rank practical principles, resolving, for instance, to prioritize an early start to the day over satisfying the inclination to sleep in, but inclinations of the moment, rather than resolved-upon rankings, determine which principles become effective.

A final, crucial component of Kant’s account is that humans are capable not only of “impure” principles of action that are “intellectual . . . in some respect,” but also of purely intellectual principles of action. Human beings have a “predisposition to the good” (6:26), a “moral predisposition” (7:324) that gives motivational force to a principle that is “purely intellectual without qualification” because it is an “impelling cause” that “is represented by the pure understanding” (28:589). A purely intellectual principle is not based in any way on one’s instincts or

inclinations but proceeds solely from practical reason itself. In chapter one, we saw the importance from the standpoint of transcendental anthropology (moral philosophy) of the possibility of an “autonomous” moral law, a principle governing human actions that does not require appeal to inclination for its justification. Within Kant’s psychology, the role of pure practical principles is different. They are principles of the higher faculty of desire that do not require positing any instincts or inclinations as factors in its explanation. In chapter two, we noted that Kant must show how moral laws can motivate human beings that are objects of empirical description. Kant’s empirical anthropology in general provides the basic biological-psychological background for such an account. Like other natural predispositions, the predisposition to the good is simply posited in human nature. Like instincts, Kant suggests that this predisposition is innate in human beings (6:27-8, 7:324), and he even offers empirical evidence for it (7:85). But like all predispositions of the faculty of desire (including instincts), experiential factors determine the extent to which the moral predisposition is living and efficacious or amounts to mere wish. Thus, for example, when one person “confronts [another] with . . . the moral law by which he ought to act . . . , this confrontation [can] make an impression on the agent, [so that] he determines his will by an Idea of reason, creates through his reason that conception of his duty which already lay previously within him, and is . . . quickened by the other . . . [to] determine himself accordingly to the moral law” (27:521). And Kant discusses various ways in which, for instance, moral education (5:155, 6:479), polite society (6:473, 7:151), and moral-religious communities (6:94ff.) can enliven one’s innate moral predisposition.

From the standpoint of the environmental and predispositional bases of moral motivation, Kant’s account of moral motivation thus fits well into his general empirical anthropology. Because of the importance of moral motivation for his transcendental anthropology, however, and especially the importance of making sense of how a finite, empirically situated being can be motivated by a pure moral law, Kant adds detailed specific accounts of the nature of the “feeling of respect” that serves as the anticipatory “pleasure” causing choice in accordance with the moral law. Kant’s account of respect for the moral law is notoriously difficult to interpret. On the one hand, Kant says that “there is indeed no feeling for this [moral] law” (5:75), but he proceeds to give a detailed analysis of the “feeling of respect for the moral law,” the “moral feeling” that is “produced solely by reason” (5:75-6).⁷⁴

⁷⁴ Understandably, readers of Kant are largely split into those that favor a “cognitivist” reading of respect –

Given that Kant posits both a transcendental and an empirical anthropology, it is understandable that he might give different accounts of the role of pleasure in moral motivation. From the standpoint of transcendental anthropology, pleasure cannot play any role in grounding decisions to act in accordance with the moral law. If I choose to do what is right because it is (or will be) pleasurable, I do not choose autonomously, and hence do not really choose to do what is right.

If the determination of the will takes place . . . by means of a feeling, of whatever kind, that has to be presupposed in order for the law to become a sufficient determining ground of the will, so that the action is not done for the sake of the law, then the action [may] contain legality indeed, but not morality. (5:71)

As an account of moral choice from the (transcendental) perspective of deliberation, this is exactly right. Whereas non-moral choices are often based, directly or indirectly, on various feelings or anticipated feelings, when one decides to do what is morally good, it should be done because the choice is morally good. And Kant rightly adds that, from this transcendental perspective, “how a law can be of itself and immediately a determine ground of the will . . . is . . . insoluble . . . and identical with . . . how a free will is possible” (5:72). But the fact that one must see oneself as free, and as bound to freely adopt the moral law as the law of one’s will, does not preclude an empirical analysis of what such a free choice “effects . . . in the mind insofar as it is an incentive” (5:72). That is, Kant can still explain how it appears when a person freely chooses to follow the moral law.

In the context of empirical anthropology, Kant could allow, consistent with his general account of human motivation, that the thought of the pure moral principle gives rise to volition by means of pleasure. In his lectures on empirical psychology, Kant even talks about an “intellectual pleasure” that arises from “representation of the [moral] law” (29:1024)⁷⁵ and that serves as the motivational transition from cognition of that moral law to action in accordance with it. Elsewhere, however, Kant worries about subsuming his account of moral motivation too closely under his account of motivation in general. In those contexts,

for whom mere cognition of the moral law, independent of feeling, motivates action in accordance with it – and those that favor a “sentimentalist” reading of respect within which the feeling of respect plays an essential motivational role. For detailed studies of these passages, see Allison 1989, Beck 1960, Grenberg 2005, McCarty 1993 and 1994, and Reath 1989.

⁷⁵ This account fits well with Kant’s insistence in the Critique of Practical Reason that “respect for the moral law is a feeling that is produced by an intellectual ground” (5:73). See too 19:185-6, R6866; 28:253-4, 674-5; 29:890, 899-900, 1013.

Kant suggests that “this singular feeling . . . cannot be compared to any pathological feeling” (5:76, cf. 20:207). In addition to its wholly intellectual ground, Kant suggests that the feeling of respect is not, strictly speaking, a feeling at all. Cognition of the moral law directly causes one to choose in accordance with it, without any specific feeling functioning as an intermediary. But “inasmuch as it moves resistance [of inclinations] out of the way, in the judgment of reason this removal of a hindrance is esteemed equivalent to a positive furthering” and so “this . . . can be called a feeling of respect for the moral law” (5:75). This account would involve modifying Kant’s empirical account of action in the context of moral actions, for which feeling would not properly speaking mediate between cognition and volition but only provide the subjective impression of cognition directly causing volitions counter to our inclinations. Whether, in the end, respect for the moral law functions just like any other feeling in mediating higher cognition (of the moral law) and volition, the fundamental structure of Kant’s empirical anthropology allows him to posit a predisposition to morality that, in conjunction with one’s character, can ground volitions that are empirically caused by the cognition of the moral law as such. Kant thus provides a framework for empirically describing what, from the standpoint of transcendental anthropology, are free choices of a morally good will.

4. Conclusion

Insofar as they are objects of empirical study, humans are biological beings with complex mental lives. As biological beings, we have various predispositions that are best discussed in terms of the purposes that they serve, and these predispositions provide grounds for causal laws that determine how our environment shapes our cognitions, feelings, and desires. The result is a complex causal account that allows for significant differences between individuals in terms of beliefs, pleasures, and choices, while still situating these differences in the context of universal laws of human psychology. So far, this account of human beings has been relatively free of moral implications. As a strictly *empirical* anthropology, there is no direct basis for ascribing moral value to any particular psychological structures over others. But Kant uses his empirical anthropology to argue for an important moral claim about human beings. As we will see in the next chapter, Kant argues that there is good empirical evidence that human beings have a predispositional structure that can rightly be called “radically *evil*.” Moreover, this chapter has emphasized humans’ empirical nature as both universal and fixed. But Kant’s empirical anthropology also includes accounts of the *historical*

change of the human species and of significant *diversity* in humans' make-up. The next chapter investigates Kant's account of human historicity, and we turn to Kant's account of human diversity in chapter five.

CHAPTER 3: HUMAN EVIL AND HUMAN HISTORY

In the last chapter, we saw that Kant has a detailed empirical account of human beings. While this account does not rise to the level of a “science” in Kant’s strict sense, it qualifies as a highly systematic account of universal human characteristics. In this chapter, we look at two further and related aspects of Kant’s empirical account of human beings. These aspects flesh out Kant’s empirical anthropology and complete the unfinished business left by the Critique of Judgment with respect to the question of what we may hope for humanity as a species (see 11:429). First, we look at Kant’s account of human evil. Kant argues that human beings are evil “by nature” and that evil is “radical” in that it affects the root of all choices. Despite this apparently glum assessment, however, Kant endorses a realistic hope for human goodness. Second, we will look at one component of this hope, Kant’s philosophy of human history, beginning with the emergence of human beings as a new kind of animal with a rational nature and moral vocation and progressing towards a future of perpetual peace amongst nations and increasingly cosmopolitan political, ethical, and social lives.

I. Radical Evil in Human Nature

a) “The Human Being is Evil by Nature”

Kant discusses human evil in his Anthropology (7:324f.) and in various lectures and notes on ethics, anthropology, and religion, but his most sustained discussion of it takes place in Religion within the Boundaries of Mere Reason, a work in which Kant aims “to make apparent the relation of religion to a human nature partly laden with good dispositions and partly with evil ones” (6:11). *Religion* starts with an argument for the existence of human evil that is complicated by what appear to be contradictory claims. At times, Kant seems to rule out the possibility of knowing anything about one’s moral status at all, saying such things as that “we can never, even by the strictest examination, completely plumb the depths of the secret incentives of our actions” (4:407) and that

A human being's inner experience of himself does not allow him so to fathom the depths of his heart as to be able to attain, through self-observation, an entirely reliable cognition of the basis of the maxims which he professes. (6:63, see too 6:36-37, 8:270)

Elsewhere, Kant suggests that even if we know that human beings are evil, we should avoid drawing attention to this evil, especially in others: "It is a duty . . . not to take malicious pleasure in exposing the faults of others . . . but rather to throw the veil of philanthropy over their faults, not merely by softening our judgments but by keeping these judgments to ourselves" (6:466). When Kant does propose arguing for human evil, he claims that it "can only be proved [by] anthropological research" (6:25) and that "the existence of this propensity to evil in human nature can be established through experiential demonstrations" (6:35, see too 6:32), but he also insists that "the judgment that an agent is an evil human being cannot reliably be based on experience" (6:20). And insofar as he does appeal to experience, Kant sometimes seems to argue directly from "the multitude of woeful examples that the experience of human deeds parades before us" (6:32-33), but elsewhere insists that his claim that "the whole species" is evil can be justified only "if it transpires from anthropological research that the grounds that justify us in attributing . . . [evil] to human beings . . . are of such a nature that there is no cause for exempting anyone from it" (6:25), which suggests that mere examples of evil, even if widespread, are insufficient. From a quick look at these passages, it becomes unclear whether there can even be an argument for human evil, and if there is, whether that argument is a priori or empirical.⁷⁶

Fortunately, things are not as hopeless as they seem, and Kant's various statements fit together into a complicated but plausible anthropological defense of human evil. The key to putting together Kant's argument comes at the very beginning of his *Religion*, in "Part One: . . . Of the Radical Evil in Human Nature":

We call a human being evil . . . not because he performs actions that are evil . . . , but because these are so constituted that they allow the inference of evil maxims in him . . . In order . . . to call a human being evil, it must be possible to infer a priori from a number of consciously evil actions, or even from a single one, an underlying evil maxim. (6:20)

⁷⁶ For various views on Kant's account of and arguments for radical evil, see Allison 1990, 2001; Frierson 2003; Grenberg 2006; Wood 2009 and 2000:287

Kant's argument for evil involves both an empirical component (the experience of "a number of evil actions") and an a priori component that justifies the inference from these to the "evil maxim" that underlies them. The rest of this section unpacks this argument and its implications for the nature of human evil.

The passage above implies that one can infer one's maxims from one's actions. While this might seem to contradict the claims quoted above about the impossibility of self-knowledge, Kant is actually remarkably consistent. Whenever Kant emphasizes the inscrutability of humans' motives, he emphasizes only that we can never know that our maxims are good.⁷⁷ But with moral evil, the case is different. While there are no actions that cannot be done from bad motives, there are some actions that cannot be done from good motives. Kant's reference, in the above quotation, to "actions that are evil" and his specification of these as "contrary to law," is important. Generally, for Kant, it is maxims rather than actions that are good or evil. But there are "actions . . . contrary to duty" (4:397), and in his *Metaphysics of Morals*, Kant articulates a political theory based on the intrinsic wrongness of actions that cannot "coexist with everyone's freedom in accordance with a universal law" (6:231). Precisely because these actions are wrong regardless of the end for which one performs them, one can legitimately infer bad underlying maxims from the performance of such actions. Moreover, because moral inscrutability comes partly from humans' tendency to self-flattery, it is implausible that one would pretend to a motive less noble than one's actual motive, so when one finds an evil motive in oneself, one can reasonably trust that there is no underlying righteous motive. Motivational inscrutability is asymmetrical: one can never know that a person – including oneself – is morally good, but one can know that people are evil.

Even if Kant's claims about inscrutability do not preclude knowledge of human evil, though, how can Kant make inferences from experience to the existence of human evil given that "the judgment that an agent is an evil human being cannot reliably be based on experience"

⁷⁷ One's moral status is inscrutable because any action that accords with the moral law might also be done from some (hidden) inclination and because human beings have a self-deceptive tendency to overestimate the moral worth of our actions. In the passages quoted above, when Kant says that "we can never . . . plumb the depths of the secret incentives of our actions" (4:407), this is a response to the fact that "We like to flatter ourselves with the false claim to a more noble motive" (4:407). And when Kant insists that a human being cannot "fathom the depths of his heart," (R 6:63), the problem is that we cannot tell whether our maxims have the "purity and stability" necessary to be morally good. (There is another, more metaphysical, reason for inscrutability. Since the maxims that are relevant to assessing moral worth are freely chosen and thus "noumenal," they can never be objects of "knowledge" in Kant's strict sense. For discussion of this point, see Frierson 2003.)

(6:20)? Neither experience nor a priori arguments alone are sufficient for Kant's proof of evil (hence both are necessary). Experience of actions contrary to duty would not be sufficient for ascribing an evil will to human beings without an argument that links those actions to evil maxims. So what is needed in order to move from evil actions to evil maxims? Given evil actions, one knows that if those actions are grounded in freely chosen maxims, then the maxims are evil. So in order to establish a connection between evil actions and evil maxims, all that is needed is an argument that human beings are free agents who choose in accordance with maxims that can ground evil actions such as those found in experience. Kant already provided much of this account in his transcendental anthropology. In his *Religion*, Kant insists that "experience can never expose the root of evil in the supreme maxim of a free power of choice, for, as intelligible deed, the maxim precedes all experience" (6:39n), but Kant's transcendental anthropology has already shown that human actions are the phenomenal expressions of noumenal, free choices; we act only on the basis of incentives that we freely incorporate into maxims of action.⁷⁸

In *Religion*, Kant adds to his general account of human transcendental freedom an account of the specific structure of the fundamental maxim that grounds evil actions. In particular, Kant makes two important additions to the account of free choice found elsewhere in his Critical philosophy. First, he argues that human choices must be grounded in a basic maxim that is either fundamentally good or fundamentally evil; no middle ground is possible. As Kant puts it,

[I]f [someone] is good in one part [of life], he has incorporated the moral law into his maxim. And were he . . . to be evil in some other part, since the moral law of compliance with duty in general is a single one and universal, the maxim relating to it would be universal yet particular at the same time: which is contradictory. (R 6:24-5)

Because morality requires unconditional and universal compliance (4:416), one who sometimes but not always acts in conformity with the moral law never really makes the moral law his ultimate motive, since any law whose application depends upon circumstances cannot be the moral law.

Second, Kant connects the account of humans' free finitude from his transcendental anthropology with the account of human predispositions in his empirical anthropology. As we saw in the last

⁷⁸ We discussed this principle in terms of the "Incorporation Thesis" in chapter one. For a discussion of the relationship between empirical claims about human motivation and the ascription of transcendental freedom to them, see Frierson 2008.

chapter, Kant's conception of a "predisposition" has wide application, covering all basic human powers and the instincts and propensities that direct the faculty of desire. In Religion, Kant employs this notion of a predisposition in his discussion of a fundamental "predisposition to the good" in human nature that consists of three distinct "elements of the determination of the human being," animality, humanity, and personality (6:26-7, cf. 7:322-5). The predisposition to animality includes basic instincts and even propensities to various inclinations that support human life. In particular, Kant emphasizes instincts for self-preservation, for propagation of the species (sexual instinct), and "for community with other humans, i.e. the social drive" (6:26). The predisposition to humanity includes our innate tendency to compare ourselves with others and thereby a propensity to "the inclination to gain worth in the opinion of others" (6:27). Finally, the predisposition to personality is "susceptibility to respect for the moral law as of itself a sufficient incentive in the power of choice" (6:27). As we saw in the last chapter, higher cognitions are generally capable of determining the power of choice, but there is a particular predisposition by virtue of which a pure higher cognition is capable of determining choice.

By subsuming the variety of human volitional predispositions under the general category of "predisposition to the good," Kant emphasizes that no natural human instincts or inclinations are evil in themselves: "the ground of evil cannot . . . be placed . . . in the sensuous nature of the human being" (6:34). But because the good predispositions of human beings include some that are not unconditionally or morally good, there is a basis in human nature for evil.

The human being (even the worst) does not repudiate the moral law . . . The law rather imposes itself on him irresistibly, because of his moral predisposition; and if no other incentive were at work in him, he would also incorporate it into his supreme maxim as sufficient determination of his power of choice . . . He is, however, also dependent upon the incentives of his sensuous nature because of his equally innocent natural predisposition, and he incorporates them too into his maxim . . . Hence the difference, whether the human being is good or evil, must not lie in the difference between the incentives that he incorporates into his maxim . . . but in their subordination . . . : which of the two he makes the condition of the other. It follows that the human being (even the best) is evil only because he reverses the moral order of his incentives in incorporating them into his maxims. (6:36, see too 6:32)

In this important passage, Kant lays out the essence of his account of human evil. Importantly, the account can be read both in terms of transcendental freedom and in terms of empirical anthropology. The transcendental reading is crucial since in the absence of a transcendental perspective on the subordination of moral to nonmoral incentives, no empirical claim can imply anything about human evil: “In freedom alone is evil” (18:212, cf. 27:295). From the perspective of freedom, when one looks at one’s action from-within, what Kant claims here is that in all choices, we have concern both for morality and for well-being (defined in terms of animal and social inclinations), but that ultimately, we subordinate one concern to the other. Our free (noumenal) nature is constituted by whether we unconditionally prioritize the moral law to nonmoral concerns or whether we allow nonmoral concerns of sufficient weight to trump the moral law. This aspect of Kant’s account depends crucially upon the account of morality from Kant’s transcendental anthropology, within which Kant shows both that human beings are transcendently free and that morality requires unconditional obedience. Here, Kant uses these claims to argue that because morality requires unconditional obedience from a transcendently free will, *any* subordination of moral to nonmoral concerns is wholly evil.⁷⁹

But Kant’s argument for human evil is not merely an argument directed towards helping readers recognize evil from-within. He also aims to make an empirical-anthropological point, that human beings are evil *by nature*. From the standpoint of empirical anthropology, the key passage above can be read as a step towards completing his empirical account of human nature. Human beings have various predispositions that can be classified in terms of animal instincts, social inclinations for recognition, and moral interests. But a complete empirical account of human beings must discern how these needs interact in cases when more than one is active. Kant thus looks for empirical evidence to suggest that the empirical character of human volition is structured such that moral grounds are inactive when they conflict with sufficiently strong nonmoral grounds. What is this empirical evidence? At the most fundamental level, Kant finds in the “multitude of woeful examples” of human misdeeds evidence for the empirical claim that humans’ volitional

⁷⁹ Kant highlights the from-within character of his argument by addressing it to individual, free agents. He emphasizes that we “notice (at least within ourselves) [actions that] are consciously contrary to [moral] law” (6:20) and gives examples of vices, especially those in civilized society, that his readers will – reluctantly – have to admit as applying to themselves. When he concludes his argument for evil in human nature, he even emphasizes this personal aspect of it. Whether or not “every man has his price, for which he sells himself” is something, Kant suggests, that “everyone can decide for himself” (6:39).

structure is such that moral incentives are subordinated to nonmoral ones. To show that this corrupt volitional structure cannot be ascribed to societal influence alone, Kant points out the presence of “unprovoked cruelty” in the state of nature, and to show that this corrupt volition is not limited to uncivilized savages, he lays out a host of vices of the “civilized state,” including “secret falsity even in the most intimate friendship . . . , a propensity to hate him to whom we are indebted . . . , many other vices yet hidden under the appearance of virtue, [and] those of which no secret is made [wherein] . . . someone already counts as good when his evil is common to a class” (6:33). This evidence gives Kant grounds for the empirical-anthropological claim that humans’ particular choices are grounded in a power of choice wherein moral predispositions are subordinated to nonmoral ones. Humans have predispositions that make evil possible and a volitional structure in which the moral predisposition is made inactive by sufficiently strong sensuous incentives. Given our transcendental freedom (established by Kant’s transcendental anthropology), human beings are thus evil. Transcendentally speaking, there is no necessity for human beings to have this volitional structure; it is contingent upon transcendently free choice. But empirically speaking, when one seeks to discern human nature based on empirical evidence, there is good reason to think that human volition subordinates *pure* higher volition to *impure* higher volition. And given that Kant’s transcendental anthropology shows this empirical character to be grounded in free choice, there is reason to describe this subordination as “evil.”

In the end, Kant’s argument for evil in human nature is simple in outline and rich in detail.

1. In widely varying circumstances, human beings perform actions that contradict the moral law and/or consciously perform actions that are immoral.
2. Human actions result from the influence of empirical causes through ordered predispositions that determine how empirical causes effect particular actions.
3. Human beings have both a moral predisposition according to which they can be motivated by the moral law and nonmoral predispositions to pursue natural and social goods.
4. The moral law is essentially unconditional, requiring stable and pure adherence.

5. Thus, human behavior is characterized by a prioritization of nonmoral predispositions over the moral predisposition.

6. Humans' empirical behavior and character express their transcendently free choices.

7. Thus, human beings are morally evil.

The first three premises are empirical generalizations, of different levels of complexity. The first is a straightforward generalization of observations about human beings. The second and third⁸⁰ generalize an anthropological explanatory model from a number of cases in which it has been observed to be a good explanation for observed human behavior. These premises are developed in much greater empirical detail, as we showed in chapter two. The fourth premise is a moral premise, a part of Kant's a priori, transcendental anthropology of volition. The evidence for this claim is thus a priori. In the context of a properly empirical argument, this a priori premise would be taken as stipulative. In that sense, the preliminary conclusion at (5) could be taken to be a properly empirical-anthropological conclusion. Given that prioritizing the moral predisposition would involve (by definition) consistency in following that predisposition, it is clear from premises (1)-(3) that human beings act according to a complex structure of predispositions within which the "moral" predisposition is subordinated to others. And in that sense, (5) expresses an empirical fact about human nature. But premise (6) is essentially transcendental; there is no empirical evidence for humans' status as free grounds of their empirical characters. Given this premise, however, the prioritization of nonmoral predispositions over the moral predisposition that was shown to be a part of human nature is also revealed as an expression of moral evil. The conclusion which is both transcendental (a priori) and empirical, is that human beings are evil by nature, that is, that moral evil can be ascribed to every member of the human species.

b) The nature of radical evil

Having shown that human beings are evil, Kant elaborates on the nature of evil. Most importantly, Kant emphasizes that human evil is

⁸⁰ In his *Critique of Practical Reason*, Kant also offers an "a priori" argument for his third premise (see 5:78). Given our obligations, we can know a priori that we have a capacity to act out of respect for the moral law, which implies that we have a *power* to do so. For the purposes of his proof in the *Religion*, however, this premise is taken as part of Kant's empirical anthropology.

“radical” in that “it corrupts . . . the subjective supreme ground of all maxims” (6:37). The “maxim” by which humans subordinate moral to nonmoral incentives is their most fundamental maxim. In general, humans act in accordance with various principles (maxims) of action, which can be ordered in a kind of hierarchy. To take one of Kant’s own examples, one might act on the maxim “when I believe myself in need of money I shall borrow money and promise to repay it, even though I know that this will never happen” (4:422), but this maxim is merely a particular application of more general maxims such as “I will trust my own assessments of my needs” and “when I can make use of others to satisfy my needs, I will do so (regardless of whether I treat them as ends when I do so),” and this latter maxim is a more specific application of an even more general maxim that Kant explains in terms of the relative subordination of inclinations and morality, something like “I will obey the moral law only insofar as doing so is compatible with satisfying other desires, and I will seek to satisfy some nonmoral desires.” This maxim, Kant suggests, is the fundamental guiding maxim of an evil human being’s life, and all other maxims are merely applications to particular cases where inclinations and/or the moral law are in play. Because this corrupt maxim lies at the root of all one’s choices, Kant refers to human beings as “radically evil.”

In laying out this account of radical evil, Kant also clarifies some important details about the nature of evil. For one thing, radical evil is not only “itself morally evil, since it must ultimately be sought in a free power of choice” (6:37), but it is also tied to a “natural propensity to evil” that structures particular evil choices that human beings make.⁸¹ Moreover, the source of radical evil in choice implies that radical evil “cannot be placed, as is commonly done, in the sensuous nature of the human being and in the natural inclinations originating from it” (6:34-5). At one level, this claim should be quite obvious. As a moral category, evil cannot properly be located in the human being qua object of empirical investigation but must be traced to the free, noumenal agent that grounds empirically observable behavior. But Kant’s claim shifts the focus even within empirical character. The empirical expression of radical evil is not in the lower faculties – the senses and inclinations – but in the higher faculties, especially in the higher faculty of desire. Human agents, even as empirically observed, have a capacity – what Kant calls “the power of choice” – to act from principles, and it is the way

⁸¹ Many commentators see this propensity to evil as a *precondition* of radical evil (cf. Allison 1990, Wood 2000), but I see Kant as portraying the propensity to evil as both a *consequence* of humans’ radical evil and as a *ground* of further evil choices (see Frierson 2003).

in which this capacity is used that gives empirical evidence of freely chosen evil.

Kant also describes three ways in which evil might express itself in one's choices: frailty, impurity, and depravity. The first involves merely a lack of character, or what Kant calls in his anthropology lectures a "bad character," an "inability to act according to principle" (25:650). In these cases, the principles of one's higher faculty of desire are good, but when it comes to acting, these principles do not actually determine one's actions. As we noted in the last chapter, there can be conflicting underlying grounds of action, and often one or more powers are "dead" or "inactive" while others are active in effecting a transition to a new mental state or an action. Those with frail wills understand the principles according to which they should act, and the character of their higher faculty of desire is such that "I incorporate the good (law) into the maxim of my power of choice, but this good . . . is subjectively the weaker (in comparison with inclination) whenever the maxim is to be followed" (6:29). In the paradigm cases of frailty, one's higher faculty of desire is properly oriented such that, if active, it would cause one to do what is right. But when the relevant moment comes, the higher faculty of desire is weaker than inclination (the lower faculty of desire) and hence inactive. We might imagine more complex cases, where inclination does not wholly overcome the higher faculty of desire but prevents it from actually bringing about effects in accordance with its character; here the inclination subverts the normal operation of the higher faculty, rather than preventing its operation altogether. In these cases, the propensity to be governed by fixed principles is not fully developed. And Religion argues that this empirically observable badness of character can be ascribed to a free (noumenal) choice to subordinate morality to inclination.

The other two forms of evil are more straightforward; both involve acting in accordance with principles of a corrupted higher faculty of desire. "Impurity" occurs when one's "maxim is good with respect to its object . . . [but] has not . . . adopted the law alone as its sufficient incentive" (6:30). One who is impure generally chooses what is morally required, but always only because it is both morally required and conducive to satisfying other desires. This conditional adherence to the moral law is, as Kant's argument for rigorism shows, no adherence at all. The final form of radical evil is "depravity," which involves a specific "propensity of the power of choice to maxims that subordinate the incentives of the moral law to others (not moral ones)" (6:30). One who is depraved might still often act in ways that seem moral, but the depraved person's power of choice is structured by a fundamental commitment to

nonmoral desires, regardless of whether these are morally permitted or not.

Importantly, Kant rejects the possibility of what he calls “diabolical” evil, the “disposition . . . to incorporate evil qua evil . . . into one’s maxim” (6:37). For Kant, even the most evil person is not motivated by evil as such. There are no immoral desires in human beings, since all desires can be traced to one form of the predisposition to the good. One is evil when one allows nonmoral desires to trump the moral law, not when one chooses evil as such. Thus Kant does not allow the possibility of cases like St. Augustine’s famous theft of pears “not to eat for ourselves, but simply to throw to the pigs[, for] our real pleasure consisted in doing something that was forbidden” (Augustine 1961: 47). For Kant, Augustine’s self-diagnosis must be mistaken; human beings do not have a desire to do what is morally forbidden *per se*. Evil arises only from putting nonmoral desires ahead of our innate moral predisposition.

Finally, it is important to note that in all of these cases, radical evil need not imply that one always chooses contrary to the moral law. To be evil is to be disposed to allow the moral law to be overridden given a sufficient sensuous incentive.

A member of the English Parliament exclaimed in the heat of debate: “Every man has his price, for which he sells himself.” If this is true (and everyone can decide for himself), if nowhere is a virtue which no level of temptation can overthrow, if whether good or evil wins us over only depends on which bids the most and affords the promptest payoff, then what the Apostle says might indeed hold true of human beings universally, “There is no distinction here, they are all under sin – there is none righteous (in the spirit of the law), no, not one.” (6:39)

Frailty, impurity, and even depravity all involve, in different ways, a subordination of the moral law to nonmoral desires. But one can be radically evil and still often do what is good, if one does what is good only because the price of doing good is, in a particular case, not too high.

c) The problem of radical evil

Kant’s claim that human beings are radically evil raises a serious problem at the intersection of transcendental and empirical anthropology, a problem that Kant spends the rest of his *Religion* trying to solve. Put simply, the problem is that evil seems inextirpable.

This evil is radical, since it corrupts the grounds of all maxims; as a natural propensity, it is also not to be extirpated through human forces, for this could happen only through good maxims – something that cannot take place if the subjective supreme ground of all maxims is presupposed to be corrupted. (6:37, cf. 6:45)

Because evil lies at the root of human choice, one cannot extirpate it through that same (evil) power of choice. One might think that being evil “by nature” precludes transcendental freedom and thus moral responsibility, but for Kant, we are evil “by nature” *through* our free choice. Radical evil is a consequence of humans’ use of their transcendental freedom; it is only because we freely choose to subordinate moral to nonmoral incentives that such a choice can be considered evil. But given that we freely choose evil *as the basis of all of our other choices*, it seems impossible to use that freedom to rid ourselves of evil.⁸² That is, it seems impossible that we can both choose evil and choose to extirpate that evil. And the problem of radical evil is made even worse by our propensity to evil. Human beings not only choose in evil ways but also deliberately cultivate both themselves and their environment (especially their social environment) in order to promote the easy exercise of evil tendencies. One whose evil manifests itself primarily in a frail will, for example, has cultivated a weakness of character that will be hard to overcome even if one tries to do so. And one who is depraved has developed patterns of self-deceptive moral justification and corrupting social interaction that will make it difficult for good intentions to fully overcome selfishness. Finally, the problem is even more acute because no matter how good one might be able to become, one has chosen badly, so one can never be a person who always chooses in accordance with the moral law:

however steadfastly a human being may have persevered in such a [good] disposition in a life conduct conformable to it, he nevertheless started from evil, and this is a debt which is impossible for him to wipe out. (6:72)⁸³

Altogether, not only is one’s choice oriented in such a way that one rejects moral reform (*radical* evil), but even if one were somehow to begin

⁸² The case here is similar to the case of “passions,” of which Kant says in his *Anthropology* that “they are incurable because the sick person does not want to be cured and flees from the dominion of principles by which alone a cure could occur” (7:266).

⁸³ There are obvious connections between Kant’s account of evil and the Christian doctrine of original sin. Although I do not read the *Religion* as merely a rational reconstruction of Christian doctrine, the similarities are undeniable. For more, see Adams 1998; Hare 1996; and Quinn 1984, 1986, and 1990.

such a process of reform, one would have to contend with self-wrought influences that make morally upright action difficult (the *propensity* to evil), and even if one somehow overcame these influences, one would never have a life that was wholly good from start to finish (one *started from evil*).

The inextirpability of radical evil *need* not pose a philosophical problem for Kant. Kant might just say that the empirical evidence shows that human beings are radically evil through their own fault, and so much the worse for us. Even Kant's claim that "ought implies can" need not be compromised by the claim of radical evil because radical evil is an evil that is self-wrought through our own freedom. Humans could be good, but we universally (but not necessarily) choose not to be; evil's "inextirpability" is *due to choice*, not a constraint on it. But Kant is committed to a moral vision that goes beyond mere insistence upon moral responsibility for wrongdoing: "In spite of the fall, the command that we ought to become better human beings still resounds unabated in our souls; consequently, we must also be capable of it" (6:45). Even if radical evil does not eliminate moral responsibility, it does seem to deny the real possibility of moral reform, which would undermine the force of the obligation to improve. In response, Kant defends moral hope, the possibility of reforming oneself morally despite one's radical evil. And this commitment to hope generates a problem: how can one reconcile moral rigorism, radical evil, and moral hope?⁸⁴

At one important level, Kant does not even try to explain how moral reform is possible given radical evil. When he points out that evil cannot be extirpated, he adds "through human forces" (R 6:37) and then suggests, "Some supernatural cooperation is also needed to [a person] becoming good or better" (R 6:44).⁸⁵ But this "supernatural cooperation" is ultimately beyond rational comprehension and even practical use.⁸⁶ The main role of such cooperation, which Kant calls "grace," is simply to reinforce the need for human beings to do their part to "make themselves antecedently worthy of receiving it" (R 6:44, cf. 6:118). Kant emphasizes that the inscrutability of grace is no greater than the inscrutability of freedom and even that humans' continuing recognition of their moral obligations reveals an enduring "germ of goodness . . . that cannot be extirpated or corrupted" (R 6:45-6; see too 6:49; SF 7:43, 58-9; A 7:322). The enduring germ of goodness shows that all people still have a capacity for goodness, and one's freedom gives an enduring but inexplicable hope

⁸⁴ For discussion of this problem in similar terms, see Hare 1996.

⁸⁵ For discussion and further references, see Frierson, 2003, pp. 114-22.

⁸⁶ E.g. R 6:117-8, 191; SF 7:43-4. For discussion, see Mariña 1997, Frierson 2003, and Adams 1998.

that this capacity can still be used well. Of course, none of these claims about inscrutability actually address the central problem of radical evil. Even grace seems inadequate, absent further explanation and especially given Kant's insistence on humans' need to be antecedently worthy of that grace.

But Kant's theoretically inadequate discussion of radical evil does highlight the proper stance to take towards the problem. Given his transcendental anthropology of cognition, Kant is surely correct that the metaphysical mechanisms by virtue of which radical evil might be overcome will never be understood by human beings. But the problem of radical evil is not, fundamentally, a metaphysical problem but a practical one. What ought one do in light of radical evil and what may one hope with respect to it? In one sense, the practical aspect of radical evil is easy to address. If evil is a free choice to subordinate the moral law to nonmoral desires, then what one must do is subordinate one's nonmoral desires to the moral law. Here Kant can do little more than exhort people to goodness and warn against self-deceptive despair or weakening of moral demands.⁸⁷ But radical evil is also a self-wrought tendency to act immorally, and it is, moreover, a tendency that is in evidence in the human by nature. And these aspects of radical evil require some account of the grounds for moral hope in the human species as a whole as well as an account of how one can work to undo and arm oneself against self-wrought evil tendencies. Kant deals with the former task in his sophisticated philosophy of human history, a history situated in the context of radical evil but one that justifies hope in humanity's future. Kant deals with the second task in his "moral anthropology," which deals with "the subjective conditions in human nature that hinder people or help them in fulfilling [moral] laws . . . , with the development, spreading, and strengthening of moral principles" (6:217). The rest of this chapter focuses on Kant's philosophy of history. Kant's moral anthropology will be discussed in chapter five.

II. Human Beings as a Historical Species

⁸⁷ To be fair, Kant's doctrine of grace can do a bit more than mere exhortation. Kant offers a detailed account of justificatory grace whereby suffering in one's life atones for the misdeeds of one's past and within which progress towards the good comes to be counted as perfect goodness. For discussion of these more specifically religious aspects of Kant's justification of moral hope, see Quinn 1984, 1990; Mariña 1997; Michaelson 1990; Frierson 2003, 2007b, and 2010b. My focus in this book will be on more specifically "anthropological" answers to the question of what one may hope, both through Kant's conception of historical progress and through his account of the moral anthropology that radical evil makes necessary and grace makes possible.

While Kant's conception of human evil draws from and leads to a historical conception of human beings, Kant is not generally known for his philosophy of history, and a historical conception of human beings can seem to be at odds with other important aspects of Kant's philosophy. Nonetheless, during the height of work on his transcendental philosophy, Kant wrote a series of papers on human history that develop his empirical anthropology through, among other things, the claim that human "predispositions . . . develop completely only in the species [and over history], but not in the individual" (8:18). The rest of this chapter lays out this historical conception of humanity.

a) Methodology

Like the anthropology discussed in the last chapter, Kant's historical methodology is primarily empirical. Kant begins his "Idea" essay by emphasizing this point:

Whatever concept one may form of the freedom of the will with a metaphysical aim, its appearances, the human actions, are determined just as much as every other natural occurrence in accordance with universal laws of nature. History, which concerns itself with the narration of these appearances, however deeply concealed their causes may be, nevertheless allows us to hope from it that if it considers the play of the freedom of the human will in the large, it can discover within it a regular course. (8:17)

But history is not "mere empirical groping without a guiding principle" (8:161), and Kant's account of predispositions provides this principle. While the empirical anthropology of the previous chapter focused on predispositions as bases of causal powers, Kant's history studies predispositions teleologically. In his Critique of Judgment, Kant argued that organic life could be interpreted via purposive predispositions: all organisms are "conceived of teleologically under the concept of a natural end" (5:376). In his writings on history, Kant adds an important presupposition of ascribing any purposive structure to an organic being:

All natural predispositions of a creature are determined sometime to develop themselves completely and purposively. . . An organ that is not to be used, an arrangement that does not attain to its end, is a contradiction in the teleological doctrine of nature. For if we depart from that principle then we no longer have a lawful nature but a purposelessly playing nature; and desolate chance takes the place of the guideline of reason. (8:18)

For most animals, this teleological assumption has implications only for the study of individual organisms. To identify a feature of an organism as a physical or behavioral predisposition, one must assume that it serves a purpose, which implies that at some point in the development of the organism, the feature will develop in the way needed to serve that purpose. For human beings, however, some predispositions are not fully realized in the life of any single person. The full development of human reason in the arts, sciences, and politics does not occur in any individual's life but happens over the history of the species. But insofar as one still treats capacities such as reason as natural predispositions, one must apply the same regulative principles to them as to other predispositions; one assumes that they will develop toward their end. And this assumption provides an "Idea" that can underlie a rationally-guided but empirically-based history that looks for ways in which humans' natural predispositions unfold over time.

b) The beginning of human history

Kant's treatment of the earliest human history is laid out in "Conjectures on the Beginning of Human History," which offers a quasi-scientific commentary on the story of humans' creation from Genesis. While some philosophers and anthropologists in the 18th century sought to show how human beings developed from other primates – the issue of the relationship between the upright posture and human reasoning was a hot topic of the day – Kant starts with "the existence of the human being . . . in his fully formed state . . . [and] in a couple" (8:110, see too 8:179). By "fully formed" Kant means only that the humans have all of their natural predispositions, not that these are all fully developed, but even this assumption means that Kant does not need to explain, as his student Herder aimed to do, how "*psychology*" arises from "determinate *physiology*," how higher cognitions arise from the contractions and expansions of "irritated little fiber[s]" (Herder 2002: 196,189). Instead, Kant starts with primitive rational and sexual beings and shows how humans developed from that stage. In this essay, the key development that inaugurates truly human history is "the first development of freedom from its original predisposition in the nature of the human being" (8:109). In his "Idea," Kant had argued that "Nature has willed that the human being should produce everything that goes beyond the mechanical arrangement of his animal existence entirely out of himself"

(8:19), and “Conjectures” shows how an animal with the mere potential for this sort of free species-development comes to have actual freedom.⁸⁸

Kant outlines four steps into actualized human freedom. Blending Genesis with Rousseau, Kant first describes how human beings come to desire objects that are not natural objects of instinct:

Instinct . . . guided the novice. It allowed him a few things for nourishment but forbade him others . . . Yet reason soon began to stir and sought through comparison of that which gratified with that which was represented to him by another sense than the one to which instinct was bound, such as the sense of sight, as similar to what previously was gratifying, to extend his knowledge of the means of nourishment beyond the limits of instinct . . . The occasion for deserting the natural drive might have been only something trivial; yet the success of the first attempt . . . was . . . decisive for his way of being. (8:111-12)

The first stage in the development of human freedom comes when humans’ cognitive faculties develop to the point at which they are capable of modifying desires. Whereas the earliest humans pursued objects to which they were instinctually drawn, at some point human beings decide to try “a [new] fruit whose outward look, by its similarity with other pleasant fruits . . ., invited him to the attempt” (8:112). Humans’ faculties of desire are no longer wholly at the mercy of their lower, sensory faculties of cognition, but become capable of control by the higher faculty of cognition, by conceptual awareness and principles for action.

Following Rousseau (and Genesis), Kant does not see this first step into freedom as wholly beneficial. The ability to generate new desires includes an ability to generate unhealthy desires; one can “concoct desires not only without a natural drive . . . but even contrary to it” (8:111). Moreover, freedom over desires causes a new problem:

Anxiety and fright . . . concerning how he, who still did not know the hidden properties and remote effects of any thing, should deal with this newly discovered faculty. He stood, as it were, on the brink of an abyss; for instead of the single objects of his desire to which instinct

⁸⁸ This freedom is still merely empirical, in that it is part of an empirical-historical account of human beings. It is a sort of capacity to act on the basis of principles (character) rather than mere inclinations. Because this empirical sort of freedom is correlated, for Kant, with transcendental freedom (see Frierson 2008), one might see this as the historical origin of transcendental freedom. Strictly speaking, however, Kant’s history of humanity, as a part of his empirical anthropology, cannot describe how *transcendental* freedom emerges.

had up to now directed him, there opened up an infinity of them, and he did not know how to relate to the choice of them. (8:112)

One motivated solely by the lower faculty of desire need barely think about proper means to one's ends. But once capable of generating new desires through reasoning, one must decide which objects are worth pursuing among an apparently infinite expanse of possibilities. But one still lacks any framework for making such determinations.

While the first stage in human freedom transformed desires in general (especially desires oriented towards personal physical needs), the second stage transforms the most intense and powerful social instinct in human beings: the sexual instinct. Following Rousseau, Kant sees a fundamental difference between the raw desire for sex and the way in which sexuality plays out in human life. Human beings overlay onto their desire for sexual gratification an interest in the beauty and even personality of the sex object. Picking up on the Biblical reference to Adam and Eve covering themselves with fig leaves, Kant envisages reason's rising influence over human desires. By covering themselves,⁸⁹ Adam and Eve make themselves more desirable, and the sexual instinct gets infused with ideals of beauty and propriety. Here reason "make[s] an inclination more inward and enduring by withdrawing its object from the senses," which "shows already the consciousness of some dominion of reason over impulse and not merely, as in the first step, a faculty for doing service to those impulses within a lesser or greater extension" (8:113).

The third step involves the "deliberate expectation of the future" (8:113), which requires still higher and more organized interactions between reason and desire. Like the first steps, the effects of this are ambivalent: it "is the most decisive mark of the human advantage of preparing himself to pursue distant ends in accordance with his vocation – but also simultaneously it is the most inexhaustible source of cares and worries which the uncertain future incites and from which all [other] animals are exempt" (8:113).

Finally, in the last stage the human being "comprehended (however obscurely) that he was the genuine end of nature" (8:114). Human beings come to see the products of nature as possible instruments for their own use (cf. Genesis 3:21), but they also recognize – albeit obscurely – that every other human being is an "equal participant in the gifts of nature"

⁸⁹ In the account in "Conjectures," Kant sees this artifice of refusal as used by both Adam and Eve. Elsewhere (cf. *Anthropology, Observations, Remarks*) Kant suggests that the art of refusal is a distinctively feminine art.

and thus can rightly make “the claim of *being himself an end*, of also being esteemed as such . . . , and of being used by no one merely as a means to other ends” (8:114). Kant does not, of course, think that the earliest human beings had worked out theories of human rights, nor that they actually treated all other human beings as equals. Kant is well aware that human beings seek to dominate each other and treat others as mere instruments to personal ends. But this domination among human beings has, according to Kant, a fundamentally different character than the struggle with the rest of nature. Among beings who are all capable of forming plans for themselves on the basis of “a faculty of choosing . . . a way of living” (8:112), influence takes a form either of blameworthy domination or of cooperation.

c) The development of human history

The emergence into freedom marks only the beginning of Kant’s historical account of humans. Before emerging into freedom, human beings were distinguished from animals only by latent predispositions to higher cognitive and volitional faculties. But upon becoming free, humans could become a truly historical species. At this stage, the importance of the claim that “human . . . predispositions . . . develop completely only in the species” (8:18) comes to the fore, and Kant adds to this a further claim central to his account of human history: “Nature has willed that the human being should . . . participate in no other happiness or perfection than that which he has procured for himself free from instinct through his own reason” (8:19). A human’s faculty of choosing for himself a way of living combined with his decisive ability to “prepar[e] himself to pursue distant ends in accordance with his vocation” generates the structure of human history, according to which all development of human predispositions occurs by humans’ own deliberate work. But Kant almost immediately adds an important caveat to this emphasis on freedom. While human history progresses by means of human choices, Nature⁹⁰ uses human choices to achieve ends that diverge from the immediate ends of the choices themselves.

The means nature employs in order to bring about the development of all [humans’] natural predispositions is . . . the unsocial sociability of human beings, i.e., their propensity to enter into society, . . . combined with a thoroughgoing resistance that constantly threatens to break up this society. (8:20)

⁹⁰ Throughout his writings on history, Kant associates Nature with “Providence” on the grounds that “supreme wisdom is required for the fulfillment of this end” (8:310).

Human beings have both a natural inclination to enter into society with others and a desire to exert superiority over others. Elsewhere, Kant explains that humans have a set of “predispositions to humanity” that “can be brought under the general title of a self-love that is physical and yet involves comparison . . . that is, only in comparison with others does one judge oneself happy” (6:27). One must be in the company of others (in order to compare oneself to them), but because “the inclination to gain . . . equal worth” fast becomes “an unjust desire to acquire superiority for oneself over others” (6:27), there arises constant strife as each seeks to assert superiority over others, even when this superiority brings no further advantage in terms of natural needs. Humans can bring themselves neither to leave the company of others nor to willingly subordinate their desires to those of their fellows.

For Kant, this “unsocial sociability” is the primary driving force of human progress: “it is this resistance that awakens all the powers of the human being [and] brings him to overcome his propensity to indolence” (8:21). Like Rousseau, Kant suggests that humans’ merely natural needs for food, rest, and sex are sufficiently limited that they do not require much development of human capacities. But the capacity to develop new desires, especially in the context of a need to prove oneself superior to others, requires that one cultivate the full range of human capabilities. “Thus happens the first true steps from crudity toward culture . . . ; thus all talents come bit by bit to be developed, taste is formed, and even, through progress in enlightenment, a beginning is made toward [forming society] into a moral whole” (8:21). At first, this might happen on a purely individual level, as human beings cultivate speed, strength, and dexterity, and then increasingly the ability to imagine and reason, along with the effort to make progress not only in sciences but in the arts. All of these steps are motivated primarily by “ambition, tyranny, and greed” (8:21, cf. 6:27), but these motives are sufficient to bring people out of indolence and into the hard work of becoming more and more perfect (though not *morally* perfect) human beings. Through humans’ unsocial sociability, nature achieves the great goal of bringing to fruition what are at first mere latent potentials for reasoning, character, scientific development, and artistic creativity.

The story does not end with individual progress, however, nor with mere cultivation of individual talents. Kant insists that the ultimate end of nature for human beings includes not merely human culture, within which human predispositions are developed, but also a form of society as a moral whole: “the greatest problem for the human species, to which nature compels him, is the achievement of a civil society administering

right” (8:22).⁹¹ Nature not only seeks to bring about individual cultivation of predispositions and the culture that follows from and facilitates that development; it also aims to bring about just relations among humans, a “society in which freedom under external laws can be encountered . . . , a perfectly just civil constitution” (8:22). This “civil constitution” has two fundamental aspects. First, it involves unions of people under “republican” forms of government.⁹² Second, it involves governments at peace with each other. True right cannot be established until all nations together affirm the principles of right, and this depends upon perpetual peace amongst nations. Human history, then, tends towards a condition within which all human societies will be organized under just, republican forms of government united into a “pacific league” of nations, a “federative union” that can “secure a condition of freedom of states conformably with the idea of the right of nations” (8:356).

Within his moral philosophy, Kant argues that just government and peace among nations are morally required ends for human beings. Given the need for freedom to express itself outwardly, we must establish conditions within which our outward expressions do not wrong others. This moral requirement gives Kant a moral reason to believe that history can progress toward such a state:

I shall . . . be allowed to assume that, since the human race is constantly advancing with respect to culture . . . , it is also to be conceived as progressing toward what is better with respect to the moral end of its existence . . . I do not need to prove this presupposition; it is up to its adversary to prove his case. For I rest my case on my innate duty . . . so to influence posterity that it becomes always better. (8:309)

Given the obligation to promote justice in human relations, one is entitled to believe that progress towards this condition is possible. But Kant’s philosophy of history also emphasizes empirical evidence that the moral interest in political right is a real force in human affairs. For example, in the response of spectators to the French Revolution,

⁹¹ The relationship between culture and a just political society shifts through Kant’s works. In the *Critique of Judgment*, a just civil society is part of the ultimate end of nature only as “the formal condition under which nature can achieve . . . the greatest development of the natural predispositions” (5:432). In the “Idea” and *Perpetual Peace*, a just civil society seems to be an ultimate goal of nature for human beings in addition to (and not merely instrumentally towards) the development of human culture.

⁹² For Kant, a “republican” form of government is one that works on behalf of the people as a whole in accordance with principles of justice. Such governments can have different structures (monarchic, aristocratic, or democratic).

the mode of thinking of the spectators . . . manifests such a universal yet disinterested sympathy for the players on one side against those on the other, even at the risk that this partiality could become very disadvantageous for them if discovered. Owing to its universality, this mode of thinking demonstrates a character of the human race at large and all at once; owing to its distinterestedness, a moral character of humanity, at least in its predisposition, a character which . . . permits people to hope for progress towards the better. (SF 7:85).

Kant does not argue that the motives of the revolutionaries themselves can be known to be good. But the universality of moral sympathy for those revolutionaries, even without any apparent cause for personal gain, provides empirical evidence of a universal moral interest (even if it does not provide evidence that this interest in fact governs most human actions). And this provides a basis for hope that human beings are still capable of genuine moral improvement.

But Kant's primary basis for hope in political progress towards more just civil society is not based on personal duty nor even on the hope that humans' moral interest in justice will outweigh their selfishness. Instead, as in the case of the development of human culture, Kant argues that humans' unsocial sociability provides grounds for progress towards more and more just institutions.

The problem of establishing a [just] state, no matter how hard it may sound, is soluble even for a nation of devils (if only they have understanding), and goes like this: Given a multitude of rational beings all of whom need universal laws for their preservation but each of whom is inclined covertly to exempt himself from them, so to order this multitude and establish their constitution that, although in their private dispositions they strive against one another, these yet so check on another that in their public conduct the result is the same as if they had no such evil dispositions . . . [W]hat the task requires one to know is how [the mechanism of nature] can be put to use in human beings in order to arrange the conflict of their unpeaceable dispositions within a people so that they themselves have to constrain one another to submit to coercive law and so bring about a condition of peace in which laws have force. (8:366)

In *Leviathan* (1660), Thomas Hobbes had shown how the unsociability of human beings – their “greed, diffidence, and pride” – gives rise to human lives that are “solitary, poor, nasty, brutish, and short,” and Hobbes argued that this state of affairs leads humans to subordinate themselves to lawgivers in order to maintain the order necessary for survival.

Following this suggestion, Kant argues that even without any moral interests, conflicts among humans will lead them to find laws to which they can subordinate themselves and others in order to achieve the peace and stability necessary for the satisfaction of their desires. Kant further argues, “even if a people were not forced by internal discord to submit to the constraint of public laws, war would still force them from without to do so” (8:365). All that is necessary is some people willing to assert their value over others by force in order to bring all into communities united under common laws. Moreover, war pushes human societies more and more towards republican forms of government (“Idea,” 8:26; PP 8:365-6), and the trials of war (combined with the “spirit of commerce”) drive nations increasingly towards “honorable peace.” As governments become more republican, the people who bear the costs of war increasingly control whether the nation goes to war and thus increasingly seek peace with other nations. And as the benefits of international commerce grow, they outweigh the benefits of war, such that nations seek a peace within which economic exchange can enrich all. Kant concludes by drawing together his empirical argument for political progress with the moral argument in such a way that the burden of the empirical argument is considerably alleviated: “In this way nature guarantees perpetual peace through the mechanism of human inclinations itself, with an assurance that is admittedly not adequate for predicting its future (theoretically), but that is still enough for practical purposes and makes it a duty to work toward this (not merely chimerical) end” (8:368).⁹³

Finally, Kant insists that political progress be supplemented by “ethical community.”

The dominion of the good principle is not otherwise attainable . . . than through the setting up and the diffusion of a society in accordance with, and for the sake of, laws of virtue . . . Just as the juridical state of nature is a state of war of every human being against every other, so too is the ethical state of nature one in which the good principle, which resides in every human being, is incessantly attacked by the evil which is found in him and in every

⁹³ While significant, neither cultural nor political progress exhaust Kant’s account of human beings as an historically-progressing species. Kant emphasizes, for example, the importance of pedagogical and educational progress towards “enlightenment,” “humans’ emergence from their self-incurred immaturity” (8:35), a progress facilitated by political rights – especially freedom of speech (8:37f.) – and also by the emergence and development of better educational institutions (2:445-452; 9:444, 448-51). Kant was, in fact, such an avid supporter of emerging new educational movements that took the striking step of collecting contributions for this project, making himself available for 3 hours a day during the period in which he was most intensely working on the *Critique of Pure Reason* (see 2:452)!

other as well . . . [This] ethical state of nature is a *public* feuding between the principles of virtue and a state of inner immorality that the natural human being ought to endeavour to leave behind as soon as possible. (R 6:94, 97)

Leaving this ethical state of nature behind requires the establishment of an *ethical* community, just as leaving behind the juridical state of nature requires a political community. But while political community is established by “external legal constraint,” ethical community depends upon mutual encouragement towards virtue; the only “constraint” applicable here is through a supposed divine lawgiver “who knows the . . . most intimate parts of the dispositions of each and everyone and . . . give[s] to each according to the worth of his action” (6:99). Even with God as “moral ruler of the world,” Kant insists that an ethical community have “purity: union under no other incentives than moral ones (cleansed of . . . superstition . . .)” (6:102). God provides no additional incentive to good actions, but merely a unified “supreme lawgiver” under whose authority members of an ethical community unite as a single “people” (6:99). As in the case of political and cultural progress, Kant suggests that progress towards this community depends upon the cooperation of nature (6:100-101) but Kant insists particularly strongly that “each must . . . conduct himself as if everything depended upon him. Only on this condition may he hope that a higher wisdom will provide the fulfillment of his well-intentioned effort” (6:101). Whereas political and even cultural progress happens through unsocial sociability, progress towards *ethical* community occurs only in conjunction with properly motivated cooperation.

For Kant, human beings are historical. Humans progressively develop innate talents and predispositions, contributing towards a culture within which arts and sciences flourish. We progress towards more just political structures, both within and among states.⁹⁴ Educational progress contributes to bringing about enlightenment, a state in which humans think for themselves. And ethical community contributes to moral development. Precisely how far this moral development goes is unclear. Given his transcendental anthropology of desire, according to which each human being is free and responsible for her own moral status, Kant seems committed to the view that fundamental moral character is an individual affair. In some of Kant’s works on human history, he emphasizes that historical progress is “not .

⁹⁴ Kant insists throughout his works that humans’ historical progress is not progress towards happiness: “only culture can be the ultimate end that one has cause to ascribe to nature in regard to the human species (not its own earthly happiness . . .)” (5:431, see too 8:20).

. . . an ever increasing quantity of *morality* in its attitudes [but only] . . . an increasing number of actions governed by duty, whatever the particular motive behind these actions may be, . . . i.e. . . . the external phenomena of man's moral nature" (CF, 7:91). Elsewhere, though, Kant suggests that historical progress does have an effect on human beings at their deepest moral level. Ethical community seems oriented towards making human beings morally good, and Kant suggests that "since the human race is continually progressing in cultural matters (in keeping with its natural purpose), it is also engaged in progressive improvement in relation to the moral end of its existence" (TP, 8:308-309).⁹⁵

One way to think about moral progress in history is in terms of the problem of radical evil. In the context of the Critique of Judgment's aim of bridging the gap between nature and freedom, Kant's reference to morally good wills as the final end of nature cannot refer simply to a noumenal will-in-itself, but must also refer to the appearance of that will in the world. The final end of nature is good human wills actually expressed in concrete human lives. And perfection of human capabilities is an important part of good wills expressing themselves. Progress in arts and sciences makes it possible for humans who aim for the happiness of others to more effectively promote that happiness, and the good will that seeks its own perfection requires a cultural context within which the resources for that pursuit are available. Moreover, given the necessity of external freedom for the full expression of one's choices, political rights are needed for good wills to fully express themselves in the world.

But radical evil poses three deeper problems for the concrete expression of goodness in human lives. Because human beings "started from evil" (6:72), the final end of nature cannot be perfect human wills but only wills that unendingly progress towards goodness. And given that radical evil involves an ongoing propensity to evil facilitated through self-deception, even this ongoing progress involves struggle against self-wrought evil tendencies. Finally, since human evil is both fundamental and rooted in the human *species*, it is not clear how one could ever begin to progress beyond one's fundamental commitment to prefer happiness to morality.

⁹⁵ A similar tension arises in Kant's Critique of Judgment, where he carefully distinguishes the "final end" of nature in human beings – the good will that is the only unconditional end – from the "ultimately end" of nature in human beings – the political and cultural progress that can be an object of empirical study (see 5:434-6), but then *connects* the ultimate and final purposes of nature, saying that "the ultimate end [is] that which nature is capable of doing in order to prepare him for what he must himself do in order to be a final end" (5:431).

Kant's account of historical progress can address at least the first two issues, and may be able to address the third. We saw in chapter one that Kant postulates immortality as a condition of the possibility of fully satisfying the moral law, but Kant's philosophy of history provides a naturalistic, secular way of understanding immortality. A human life can be considered a good life as a whole insofar as it not only gradually improves in its own individual pursuit of virtue but also works towards an unending progress in the expression of morally good deeds through reforming the society of which it is a part. The historicity of human nature makes it possible for one's own struggle against evil to be part of an enduring struggle of humanity as a species. In particular, and this aligns the first issue with the second, part of one's struggle against radical evil involves enacting social conditions that work to strengthen virtue rather than one's evil propensities. The nature of self-deception is such that one seeks both to excuse misdeeds on the grounds of incapacity or comparison with others and to ignore the moral law itself. Progress in culture, for Kant, involves as a central component the "culture of discipline," that "consists in the liberation of the will from the despotism of desires" (5:432). This progress directly combats the frailty of will by virtue of which one lets one's good intentions be overwhelmed by inclinations. And the "ethical community" is a community of people constantly reminding one another of their moral obligations, holding one another accountable in ways that, without being judgmental, makes it increasingly difficult to ignore the demands of morality in self-deceptive ways. In the context of human beings as initially radically evil but potentially in revolution against that evil, even not-strictly-moral cultural and political progress can profoundly affect the extent to which one's revolution expresses itself in concrete improvements. Those whose fundamental moral disposition is one of struggle against evil might, in early phases of human history, be largely dominated by evil tendencies and show only the slightest glimmers of success in the struggle against it, while those at later stages of historical progress, being increasingly armed against the evil principle through social structures that facilitate morality, will express their good wills more and more fully in their concrete, embodied lives.

These sorts of moral progress in history still leave open the question of whether historical progress can go all the way down, actually enabling or facilitating the revolution in fundamental maxims. And here one might take a clue from Kant's discussion of supernatural influence. Just as "the concept of a divine concursus is quite appropriate and even necessary" "so that we should never slacken in our striving towards the good" (8:362), but we should not use appeals to divine cooperation to

excuse moral complacency; so we might appeal to moral progress in history as encouragement that our struggle against evil will bear real fruit, but must appeal to this progress only in such a way that it prevents rather than justifies complacency. Kant's philosophy of history can thereby provide empirical support for the moral hope that is justified religiously by appeal to God's grace and our immortality.⁹⁶

III. Conclusion

In chapter one, we saw how Kant's transcendental anthropologies of volition and feeling contribute to answering the question "What may I hope?" through the postulates of God and immortality and through the recognition of human beings as ultimate and final end of nature. But when Kant introduced his questions, he associated "What may I hope?" with *religion* and claimed that *Religion within the Boundaries of Mere Reason* is where he tried to answer that question (11:429). While Kant's transcendental anthropology provides an overall framework within which hope can be justified, it is only in his religion and history that this framework is given an empirical content. While the empirical anthropology laid out in the last chapter primarily fills in the empirical account of human beings for which Kant's *Critique of Pure Reason* paved the way, the empirical studies of human evil and history are needed to complete his transcendental philosophy as a whole by providing assurance that the empirical world is conformable to the moral demands of freedom for radically evil beings like us.

⁹⁶ For further discussion of the possibility of moral progress in history, see Anderson-Gold 2001, Cohen 2010, Frierson 2003, Kleingeld 1995, Loudon 2000, Wood 1999, and Yovel 1980.

CHAPTER 4: KANT ON HUMAN DIVERSITY

Much of Kant's anthropology emphasizes universality and uniformity. His *transcendental* anthropology implies that there are proper ways of cognizing, acting in, and even feeling about the world that are universally applicable to all people. Even Kant's empirical anthropology describes general properties of human nature; while Kant recognizes that "circumstances of place and time . . . produce *habits* which, as is said, are second nature," he insists that anthropology should aim to overcome this "difficulty" in order to "rise to the rank of a formal science" (7:121). And Kant's claim that "the human being is evil by nature" is supposed to be based on "anthropological research that . . . justif[ies] us in attributing . . . [evil] to human beings" in such a way that "there is no cause for exempting anyone from it" (6:25).

Throughout his life, however, Kant was also preoccupied with human differences. Kant lectured more on "physical geography" than any other subject, and especially during its early years, this course included substantial attention to cataloging differences between different types of human beings. He describes the content of this course in 1765, saying "The comparison of human beings with each other, and the comparison of the human being today with the moral state of the human being in earlier times, furnishes us with a comprehensive map of the human species" (2:312-3). Moreover, from the start of his anthropology course in 1772, Kant included discussion of differences between human beings based on variations in temperament, nationality/ethnicity, and sex. In his published *Anthropology*, Kant emphasizes "an advantage for the reading public" in offering "headings under which this or that observed human quality . . . can be subsumed" and thus offering "readers many occasions and invitations to make each particular into a theme of its own, so as to place it in the appropriate category" (7:121-2). Among these "headings" one finds classifications of different sorts of talents and inclinations, mental illnesses, temperaments, and ethnic and gender differences. Thus even while unifying human diversity under universal

normative and descriptive headings, Kant also distinguishes between human beings, which helps make his anthropology both more “popular” (7:122) and more practically relevant. Given his aim, in his anthropology and physical geography, of “making good [the] lack of experience” of his pupils, one might helpfully compare many of Kant accounts of human diversity to popular contemporary ways of making people more adept at dealing with one another through the use of personality profiles such as the Myers-Briggs test, popular books on gender differences, or programs fostering intercultural awareness.

This chapter focuses on Kant’s account of human variation. I start with a brief treatment of Kant’s accounts of individual differences, including ordinary variations between people in terms of talents, inclinations, and natural aptitudes as well as those extraordinary variations that Kant classifies as mental disorders. I then turn to human temperaments, the four basic affective-volitional structures into which every human being can be classified. Finally, I turn to two of the most controversial aspects of Kant’s account of human beings, his discussions of sexual and racial/ethnic difference. Most readers of Kant (rightly) see him as deeply misogynistic, though precisely in what *way* he is misogynistic is complicated and controversial. Likewise, Kant’s comments on other races – especially Black Africans – are shameful, but precisely how these comments relate to his more systematic concerns – including his systematic race theory – is far from straightforward.

I. Individual Variations

Within Kant’s empirical anthropology, human beings are unique in their particular configurations of predispositions and powers. Chapter two noted that human beings have universal, natural predispositions that govern cognition, feeling, and desire, but the precise way in which these predispositions unfold is not universal. Many differences between individuals are ascribable to environmental differences, such as why one person plays cricket while another plays baseball or why individuals have different beliefs and tastes. But other differences are, to varying degrees, innate.

The most extreme individual differences are found in Kant’s accounts of mental disorders.⁹⁷ For Kant, mental disorders affect each of the three fundamental human psychological faculties: cognition, feeling, and desire. Because cognition is sub-divided into different powers

⁹⁷ For detailed discussion, see Frierson 2009a and 2009b and Shell 1996:368-305.

(imagination, judgment, etc.), Kant distinguishes cognitive disorders according to power is affected and how. For example, dementia (*Wahnsinn*) is “deranged” imagination, while craziness (*Aberwitz*) is deranged reason. Kant distinguishes between mere deficiencies and positive forms of derangement, such that, for instance, stupidity is a deficiency of judgment whereby one simply lacks the ability to figure out whether a particular case falls under a general rule, while insanity (*Wahnwitz*) is a derangement of judgment whereby one groups together disparate particulars under false universals. Deficiencies involve lacking a particular power or having a power that is so weak as to be largely inactive. Derangement takes place when a cognitive power is governed by laws different from those of other human beings. Whereas judgment normally groups particulars under universals in accordance with certain real similarities, one who is insane connects particulars in ways without comprehensible bases. (Kant suggests that those with deranged judgment can often be entertaining poets, precisely because they link together particulars in bizarre but quasi-systematic ways.) Further, Kant adds melancholia and hypochondria as cognitive disorders distinct from those that fall under more general groupings. Regarding feeling and desire, Kant treats all disorders of feeling under the general name of “affects” and disorders of desire under the general name of “passions.” Both are states wherein a particular feeling or desire overpowers the reflection needed to compare that feeling or desire with others, so a single feeling or desire motivates action without (sufficient) reflection. Finally, Kant describes origins of mental disorders and ways of treating them. Madness is ascribed to a biologically-inherited “germ” that sets on at a particular time and takes on its particular character due to circumstances present when it sets on (7:217). Hypochondria results from a “natural predisposition” (7:104) that has the form of a propensity and can be resisted through “intentional *abstraction*, [which] may weaken the feeling[s on which the hypochondriac dwells], and if the abstraction becomes habitual, make it stay away completely” (7: 212).

While interested in “bringing a systematic division” into mental disorder, Kant also classifies differences between ordinary, mentally-healthy human beings. Such people have the same mental powers that operate by the same general rules, but there is still room for difference in the details of their operation and the relative weight of different influences on thought and action. Kant classifies these human differences into two general categories, those that “indicate what can be made of the human being” and those that “indicate what he is prepared to make of himself” (7:285). The former can be further sub-divided into

talents, natural aptitudes, and temperament.⁹⁸ The latter is identified by Kant with “character purely and simply” (7:285) and has been discussed in chapter two. While character is an important source of human differences, the biological basis of these differences is the universal propensity to character; what determines whether and how this propensity manifests itself in particular human beings is largely⁹⁹ a matter of non-biological influences on one’s life, including “what one . . . make[s] of oneself.” By contrast, talents, natural aptitudes, and temperaments are variations in natural predispositions, or at least variations “founded upon . . . [different] natural predisposition[s]” (7:286, see too 7:220). None are distinct mental powers in themselves; rather, they describe the degree to which various natural powers are capable of being exercised or improved. Talents refer to “excellence[s] of the cognitive faculty” (7:220), natural aptitude “has more to do . . . with feeling” (7:286),¹⁰⁰ and temperament “has . . . to do . . . with the *faculty of desire*” (7:286). Moreover, all of these natural variations “must . . . be distinguished from . . . habitual disposition (incurred through habit) because a habitual disposition is not founded upon any natural predisposition but on mere occasional causes” (7:286). Beyond humans’ shared mental powers and the differences acquired through different lives and experiences, there are also innate differences in the degrees to which and ways in which mental powers can be exercised.

Talents, for Kant, always indicate an especially effective use of one’s cognitive faculties. The talents Kant emphasizes in his *Anthropology* are “*wit . . . , sagacity, and originality of thought (genius)*”

⁹⁸ Kant’s lectures on anthropology show significant shifts in his conceptions of talents, natural aptitudes, and temperament from his first interest in these topics (at least as early as 1764 in his *Observations*) through the published *Anthropology*. With the exception of my discussion of his shift from preferring the melancholic temperament to preferring the phlegmatic, I gloss over those changes here, focusing on *Anthropology*.

⁹⁹ There are important biological factors – such as temperament, sex, and race – that can have an influence on how one’s character develops.

¹⁰⁰ Kant’s discussions of natural aptitudes (*Naturell*) are extremely abbreviated, not only in his *Anthropology* but throughout his corpus and even his anthropology lectures, and while Kant often distinguishes natural aptitudes from talents and temperaments, he does not consistently do so, nor does he always do so in the same way. Often, Kant refers to natural aptitudes as “powers of mind” (*Gemütskräfte*, see e.g. 25:556), but in the *Anthropology*, this is narrowed to refer specifically to those natural tendencies that have “to do with the *feeling* of pleasure or displeasure, as to how one human being is affected by another (and in this natural aptitude can have something characteristic)” (7:286). With respect to both talents and natural aptitudes, Kant’s emphasis is on the range of purely individual variation that is possible when it comes to the degree to which and the ways in which our natural cognitive and affective powers can be exercised.

(7:220). The first is an excellence of understanding, whereby one can “assimilate heterogeneous representations that . . . lie far apart from each other” (7:220) in a way that is entertaining and agreeable. Sagacity is a “natural gift for judging in advance,” whereby one’s reflective judgment is effective in “knowing how to search well” and “discover[ing] things” (7:223). And genius is a talent “of imagination” (7:224) whereby one can “*invent*” things. Within Kant’s anthropology, talents serve to distinguish different sorts of people, and knowing what talents someone has (or lacks) is essential to knowing how properly to educate them (since talents themselves cannot be taught, but only cultivated). In addition to helping distinguish between people for purely anthropological purposes, however, Kant’s conception of talents plays important roles in his moral philosophy and aesthetics. Within moral philosophy, Kant argues that human beings have a special obligation to cultivate their talents. Readers who focus exclusively on Kant’s moral writings might read this to refer to a general obligation on the part of human beings to improve themselves, when Kant’s argument is more specific and crucially grants dignity (and even providential arrangement) to human diversity. Because human beings have different talents, they require different sorts of self-cultivation. These innate, biological-anthropological differences are also important for aesthetics, where Kant argues that artistic “genius” – that capacity for “giving the rule to art” and “producing that for which no rule could be given” – is a “talent” and even a “natural predisposition” (5:307, cf. 5:317-8). He thereby argues that where other accomplishments must be ascribed to human beings themselves (8:19), fine art is the work of Nature itself: “genius is the inborn predisposition . . . *through which* nature gives the rule to art” (5:307).

II. Temperament

Amongst the natural variations that constitute “what nature makes of the human being,” by far the most important is temperament. Whereas talents and natural aptitudes are highly individual, Kant holds that one can classify people into precisely four “temperaments.” Like talents and natural aptitudes, temperaments do not refer to specific additional powers or faculties, but to systematic ways in which humans’ powers differ.

Kant’s discussion of the temperaments takes place in the context of a long tradition of treating human beings as divisible into the four categories of sanguine, choleric, melancholic, and phlegmatic. These categories were originally developed within medicine by Hippocrates and

Galen, who referred to ways in which bodily “humors” (phlegm, yellow and black bile, and blood) were mixed in the body. The idea was that bodily health required perfect balance, and one’s “temperament” was the particular way in which one’s humors were unbalanced (so, for instance, too much blood made one sanguine; too much phlegm made one phlegmatic). By the 18th century, the theory of temperaments had become an important part of moral psychology, and 18th century moralists and proto-psychologists used “temperaments” to classify different moral characters. In line with this tradition, Kant distinguishes temperaments in terms of the different tendencies of action and motivation that can come to have moral (and, more generally, pragmatic) importance.

Kant brackets medical and physiological conceptions of temperament (see 7:286), while still acknowledging that “temperaments which we attribute merely to the soul may well also have corporeal factors in the human being, as covertly contributing causes” (7:286). Kant divides the four temperaments into sanguine and melancholic “temperaments of feeling” and choleric and phlegmatic “temperaments of activity” (7:286-7, 289).¹⁰¹ The sanguine “is carefree and of good cheer” and lives in the moment (7:288). The melancholic is serious, thoughtful, and tends towards misanthropy (7:288). The choleric is “*hot-tempered . . . [and] rash [and] his ruling passion is ambition*” (7:289). Just as the sanguine *feels* quickly and easily but is also quickly distracted, so the choleric *acts* quickly but is quickly appeased. As the sanguine is has an excess of (cheerful) feeling, the choleric has an excess of activity. Finally, the phlegmatic has “*lack of emotion*” and “the quality of not being moved easily” (7:289-90).

Although generally dismissed today, Kant’s discussion of temperament is important for several reasons. First, it is an important part of completing Kant’s empirical anthropology. In the absence of some accounts of temperaments, Kant might rightly be accused of failing to recognize the important natural (even biological) differences between human beings. Kant’s doctrine of temperaments, like more contemporary psychological investigations and classifications of human psychological variations, provides his universal and historical anthropology with a necessary supplementary account of human difference. Moreover, as one important commentator has noted, Kant’s account of human difference helps flesh out his otherwise simplistic accounts of such things as

¹⁰¹ Although temperament is distinguished from natural aptitude in that it relates to the faculty of desire, temperaments relate to the faculty of desire either by being *directly* active or by influencing feeling *in a way that is relevant to activity*.

“unsocial sociability,” which we can now see “describes a more complex set of interactions than the mere push and pull of misanthropic egoism and philanthropy. The several vices which propel history . . . are arguably attached in Kant’s mind to different kinds of people” (Larrimore 2001: 285). Temperamental differences affect how our inclinations and even character develop and thus have profound impacts on how we behave and relate to one another. Moreover, within the emerging empirical-psychological discussions of the temperaments, Kant’s way of defending temperaments makes use of (and thus partly helps justify) other aspects of Kant’s empirical and even transcendental anthropology. The neat classification of the four temperaments in terms of strength or weakness of feeling and desire helps vindicate both the legitimacy of temperament theory in terms of Kant’s innovative distinction between feeling and desire, and the validity of that distinction in terms of its ability to account for temperaments. And Kant’s treatment of the phlegmatic, as we will see shortly, both supports and is supported by Kant’s work on the nature of moral motivation.

Second, Kant’s account of temperaments is important as part of a specifically *pragmatic* anthropology. I discuss pragmatic anthropology in more detail in the next chapter, but here it is important to note that Kant does not *merely* classify different temperaments. He also emphasizes their characteristics in ways relevant to moral and practical assessments and deliberations. For example, when Kant claims that the sanguine person “makes promises in all honesty, but does not keep his word because he has not reflected deeply enough beforehand” (7:288), his advice not only provides needed warning to the sanguine about their own morally pernicious tendencies but also helps others know how to deal with sanguine companions and even how to properly evaluate the moral status of the sanguine’s broken promises (as flightiness, not deception). Similarly, when the melancholic is cautious about making promises (7:288) and the phlegmatic “proceed[s] from principles” (7:290), Kant’s ascription of these traits to temperament and *not* to “moral causes” (7:288) or “wisdom” (7:290) is an important warning to all concerned not to mistake mere temperament for genuine moral worth.¹⁰²

¹⁰² In this context, Kant’s note at 15:758-65 on “mistakes” or “failings” (*Fehler*) associated with different temperaments is an excellent example of working through the challenges that those with different temperaments will particularly have to face in living and good and happy life. My reading of Kant here differs importantly from that recently offered by Mark Larrimore. I disagree with Larrimore’s claims that a phlegmatic temperament “is a duty” (Larrimore 2001: 284) and that phlegma should be identified with apathy. Phlegma is a temperament that is *like* apathy and that makes it particularly easy to develop true moral character (which includes apathy), but I take Kant’s repeated claim that phlegma can do what philosophy or wisdom does *without real philosophy or wisdom* to be an important *warning* to the

Third, Kant's discussion of temperaments provides one of the best examples of how developments in his transcendental anthropology intersect with those in his empirical (and pragmatic) anthropology. The dominant traditional view of temperaments portrayed sanguine and choleric as healthy, phlegmatic and melancholic as inherently unhealthy (Larrimore 2001: 264). Kant consistently resisted this tradition. In his early writings, Kant emphasized, alongside a major counter-tradition in German (especially Pietist) thought, the merits of the melancholic temperament: "Genuine virtue from principles therefore has something about it that seems to agree most with the **melancholic** frame of mind in a moderate sense" (2:219). At the time that Kant wrote this, his moral philosophy was largely informed by sentimentalist views that emphasized depth of feeling as an important part of virtue. But as Kant's moral theory moved towards the importance of a purely rational moral principle and a distinctive moral motive, and as his general empirical anthropology came to emphasize the importance of "character" in human life, Kant's account of the temperaments shifted to privilege the phlegmatic over the melancholic. In his early writings, Kant followed the traditional view that the phlegmatic was hardly worth discussing (2:224). But by 1778, Kant already associated the phlegmatic with "the philosopher" (25:1167) and had developed an important distinction between what he would later call "phlegma as weakness," which is a mere "propensity to inactivity" and "phlegma as strength," or "apathy," which is a "quality of not being moved easily or *rashly*" and associated with acting "from principles and not from instinct" (7:290).

Finally, Kant's discussion of temperament provides an important "hinge" (Larrimore 2001:270) between Kant's general and universal anthropology and his discussions of differences between human sexes, nationalities (or ethnic groups), and races. Especially in his early works, Kant associates different sexes, nationalities, or races with different temperaments, and he uses claims about the latter to explicate his claims about the former. Given the offensive nature of Kant's views about sex and race, this calls for thinking about what, if anything,

phlegmatic. Whereas a sanguine person who lives life according to principles can and should regard that as a moral accomplishment, the phlegmatic one should not. Instead, the phlegmatic person should be grateful to nature or God for that temperament, but work all the harder to ensure that her principled action flows from a steady moral character and not merely a fortunate temperament. I also see no reason to think that, for Kant, one can *change* one's temperament, though one can certainly learn not to manifest its (negative) characteristics. Unlike character as such, temperament is, after all, something that "nature makes of the human being," not something that "the human being makes of himself." (This is not to say that one's temperament cannot change – Kant associates different temperaments with different ages (see 25:820) – but only that one's temperament at a give time is something that arises naturally, not through one's own effort or decision.)

distinguishes Kant's practice of subdividing people according to "temperament" – shared, for example, by those who favor psychological personality tests as a way of improving interpersonal relationships – and his practice of subdividing people according to sex and race. Given that for Kant temperaments are as much a matter of what "nature makes of the human being" as sex or race, why is one sort of distinction offensive while the other might seem at worst merely naïve? Or put another way, given the dangers of distinguishing on the basis of race and sex, might it not be appropriate to be concerned about *any* simplistic attempts to classify different sorts of people?

III. Differences between the Sexes¹⁰³

Kant's discussion of temperament marks the start of his attempt to classify human beings not only in terms of universally-shared characteristics or individual variations but in terms of generic *types* of human. But throughout his anthropological writings, beginning from his very early *Observations on the Feeling of the Beautiful and Sublime* (arguably his first foray into anthropology) and continuing through his published *Anthropology*, Kant follows his discussion of human temperaments with an account of the difference between men and women, a discussion, that is, of "The Character of the Sexes" (7:303). Consistently, and in perfect conformity to feminist characterizations of Western discourse as fundamentally patriarchal, Kant's discussion of differences between men and women focuses exclusively on the unique character of *women*. Kant takes men to be paradigmatic of human beings in general, such that a characterization of "the sexes" involves nothing more than showing how the previous characterization of human beings in general must be modified for the "special case" of women.

Kant's *Observations*, his earliest (1764) and most popular (including amongst women!) anthropological discussion of the sexes, includes both perfect sound-bites of Kantian misogyny – "A woman who has a head full of Greek . . . might as well have a beard" (2:229) and apparent mantras of egalitarianism – "the fair sex has just as much understanding as the male" (2:229). The core of Kant's account of the sexes in *Observations* is that women are primarily beautiful, while men are primarily sublime:

¹⁰³ Throughout his discussion of differences between men and women, Kant conflates what we would now call sex-differences and gender-differences. He also assumes heterosexuality throughout his discussions of relations between sexes.

it is not to be understood that woman is lacking noble [sublime] qualities or that the male sex must entirely forego beauties; rather one expects that each sex will unite both, but in such a way that in a woman all other merits should only be united so as to emphasize the character of the **beautiful**, which is the proper point of reference, while by contrast among the male qualities the **sublime** should clearly stand out as the criterion of his kind (2:228).

Kant's distinction is both descriptive – women *are* generally more characterized by the beautiful and men by the sublime – and normative: “To this [distinction] must refer all judgments of these two sexes, those of praise as well as those of blame” (2:228) such that “what is most important is that the man become more perfect as a man and the woman as a woman” (2:242-3).

Unless one keeps both the descriptive and normative dimensions of Kant's distinction in mind, Kant's account might seem to preclude virtue in women. Kant says both “It is difficult for me to believe that the fair sex is capable of principles” (2:232, see too 27:49), and “true virtue can only be grafted upon principles” (2:217). This might require, as Jean Rumsey claims, that “women . . . are in Kant's view less than . . . full moral agents.”¹⁰⁴ But such attention to the merely descriptive aspect of Kant's distinction misses Kant's insistence in *Observations* that women *are* capable of virtue, but “The virtue of the woman is a **beautiful virtue**” (2:231, see too 27:49-50). Following through on his sexual distinction, Kant insists that women are capable of distinctively feminine virtue. And whereas the principles of which women are incapable “are also extremely rare among the male sex” (2:232), the “love [of] what is good” that serves as the foundation of beautiful virtue is grounded in “goodly and benevolent sentiments” that “providence has implanted . . . in [woman's] bosom” (2:232). The impossibility of fulfilling *male* virtue is actually a moral advantage; whereas few men will attain sublime virtue, women are well equipped for beautiful virtue.

By the time of Kant's *Anthropology*,¹⁰⁵ however, Kant's thought underwent several changes that affect his discussion of women. Some of these reflect Kant's increasing interest in courtship and marriage. So, for

¹⁰⁴ Jean Rumsey, “Re-Visions of Agency in Kant's Moral Theory,” in *Feminist Interpretations of Immanuel Kant* (Ed. Robin May Schott), University Park, PA: The Pennsylvania State University Press, 1997, p. 131.

¹⁰⁵ Many of the additions are already reflected in *Remarks* written in 1764-6 in Kant's personal copy of his *Observations*. The mid 1760's, when Kant was personally deciding whether or not to marry (he eventually decided against it) and reading the works of Rousseau (who's *Emile* and *Julie* extensively discuss differences between the sexes) seem to have been the period in which Kant did the most reflection on the differences between men and women, and during which he formed most of his views on this matter.

example, Kant's personal notes and lectures in anthropology increasingly emphasize that it is "an essential condition of nature, that woman must be sought" (25:708), so that "the woman *refuses*, the man *woos*; her surrender is a favor" (7:306). This characterization of "natural" courtship practices contributes to Kant's attention to a womanly "art of appearing" or "art of illusion" (20:61, 69, 121, 140). Since women in fact want (and need) men as much as men do women, their refusal is a ruse, a way of attracting men, and this art of illusion continues throughout marriage as a way to "govern . . . men and use them for their own purposes" (7:304). The woman "acquires confidence early in her ability to please" (7:306), especially through presenting charming appearances, and for this reason, women *care* much more about appearances than men (25:720).

More important than these details of his account, though, Kant reconceived of the difference between the sexes in terms of an overall natural teleology. From his earliest discussions of women, Kant had referred to their "innate" characteristics (2:229) and insisted that women not only are but ought to be different from men (2:229-30). But his later anthropology takes this further. First, Kant clarifies the extent to which these differences are natural. He recognizes that there are substantial differences between 18th century European relations between the sexes and those in what Kant calls "uncivilized conditions." In such conditions, Kant argues, one cannot see any major differences between men and women, except that men are somewhat physically stronger (and more "courageous"). But, Kant claims, this greater sameness between men and women is hardly a boon to women, who find themselves, without distinctively feminine sources of strength, in conditions of "domestic animal[s]" (7:304). Consistent with his account of predispositions in general, Kant sees feminine character traits as propensities that require the right conditions to flourish: "culture does not introduce these feminine qualities, it only allows them to develop and become recognizable under favorable conditions" (7:303). Kant also offers specific arguments against those who "dispute this [account of sex differences] in the way one disputes something from the speaker's lectern [to show that they are not inherent to] nature and [they] believe it to be a matter of fashion" (25:709). Kant appeals to "universal and constant" facts about the sexes, including not only that women bear children while men do not, but even that for humans as well as "animals . . . , one sees the female is the refusing, but the male the courting party" (25:709). From such universal characteristics, Kant argues, the further characteristics of human females – their abilities to please through illusion, a desire to dominate men through charm, etc – can be explained.

The most important part of Kant's account of natural differences between men and women is his treatment of "nature's end in establishing womankind" (7:305). There are two main natural purposes for women's distinctive characteristics: the preservation of the species [and] the cultivation and refinement of society (7:305-6).¹⁰⁶ Given that "nature entrusted to woman's womb its dearest pledge, namely, the species, in the fetus," women's "fear of physical danger" is an important insurance towards her preservation (7:306). Combining this physical "weakness" with an ability to "demand male protection" ensures that the fetus (and thus the species) will be threatened neither by excessive boldness on the part of the woman nor physical dangers (against which the man will protect her). The second great end of nature is the cultivation of society: "nature wanted to instill the finer feelings that belong to culture – namely, . . . sociability and propriety" (7:306). As we saw in the previous chapter, the development of culture is the great natural end for human

¹⁰⁶ There is a third, more immediate, purpose, but Kant does not discuss this as a purpose *of nature*. In the immediate domestic sphere, women's capacity for illusion and for "mastering [man's] desire for her" (7:303) ensures a level of stability and equality within marriage.

Two persons convening at random is insufficient for the unity and indissolubility of a union; one partner must *yield* to the other, and, in turn, one must be superior to the other in some way, in order to be able to rule over or govern him. For in the *equality* of claims of two people who cannot do without each other, self-love produces nothing but squabbling. (7:303)

While Kant's approach to marriage can seem extremely patriarchal (as we will see shortly), this passage suggests something more complicated. For Kant, a lasting relationship between two people, especially when sexual, is unsustainable without some sort of hierarchy. But Kant suggests that apparent hierarchies are not (and should not be) entirely what they seem. Whereas the claim that "one partner must yield" and another "be superior" might easily be taken to justify male superiority within the marriage union, Kant in fact uses it to argue for the claim that "each partner must be superior in a different way" (7:303). The man is superior in obvious ways, through "physical power and courage" and – in society – through formal political rights. But Kant suggests that the *real* superiority is in the hands of the woman. Through her charms, "domestic warfare which she conducts with her tongue," and carefully used "tears of exasperation" (7:304), she is able to govern and control the men who, it turns out, only *seem* to rule over her. Woman's art of appearances thus serves the purpose of balancing out men's greater strength and thereby promoting equality within marriage (and, more generally, within male-female relationships).

Kant even worries that in his age, women have actually gone beyond equality and exercise superiority over men. He contrasts the state of polygamy, where men exercise undue control over women, the civil condition of marital equality, and the state of luxury. In the last, "gallantry has become the fashion and jealousy ridiculous. . . , [so] the feminine character reveals itself: by extending favors toward men, woman lays claim to freedom and, at the same time, to the conquest of the entire male sex" (7:304-5). The new polygamy of the "modern" (18th century) age, Kant suggests, is one within which "the time of the debaucheries of men has ended and that of women has begun" (20:86).

beings as a species. There Kant emphasized unsocial sociability as the driving force behind this development. Here he highlights that unsocial sociability has a gendered structure. Men and women are attracted into society with one another but manifest their superiority in different ways. In particular, women's power over men depends upon increasingly polite and refined social interactions; her direct power is exercised through "modesty and eloquence in speech and expression" (7:306). In order to gain equality, women become adept at social interaction. But as women become more capable of coaxing men, they "claim . . . gentle and courteous treatment by the male," who finds himself "fettered . . . through his own magnanimity, and led by her, if not to morality itself, to that which is its cloak, moral decency" (7:306, see too 2:241). The apparent weakness and timidity of women ends up becoming one of the driving forces behind cultural and even proto-moral progress in the human species.

Unfortunately, Kant's increased interest in women as a driving force behind progress in history, even to the point of helping develop a moral decency "which is the preparation for morality" (7:306), was accompanied by profound changes in his overall moral theory, changes that effectively precluded women from being capable of virtue. Whereas the Kantian ethics of *Observations* emphasizes the importance of "beautiful" or "adopted" virtues even for men and devotes significant attention to spelling out the details of "the virtue of women" (2:231), Kant's mature moral theory not only does not include, but even seems to preclude, anything that could *genuinely* be called feminine *virtue*. Kant's shift from an empirical and sentimentalist moral theory in the 1760s that allowed different sorts of moral worth based on different aesthetic feelings to a more rigorous rationalist morals in his *Groundwork* and later works that emphasizes a "good will" as the only thing "good without limitation" (4:393) precludes taking seriously as "virtue" anything that does not involve acting out of respect for a pure moral law. Thus when he refers to "feminine virtue" (7:307) in his *Anthropology*, the claim seems to be a mere remnant of an earlier view, a remnant that no longer makes sense in the context of Kant's mature moral theory.

At the same time, unfortunately, Kant's anthropological characterization of women as incapable of male virtue (which becomes the only real virtue) is unchanged. The early claim that "It is difficult for me to believe that the fair sex is capable of principles" (2:232)¹⁰⁷ blossoms into a more technical (and more problematic) claim that certain

¹⁰⁷ Kant immediately adds, "and I hope not to give offense by this," in keeping with the gallant tone of *Observations*.

“feminine principle[s are] hard to unite with a *character* in the narrow sense of the term.”¹⁰⁸ This “narrow sense” of character is the capacity to act on consistent principles of one’s own, a capacity that not only “has an inner worth” of its own (7:293) but is also a necessary condition of a good will (which makes the moral law its principle).¹⁰⁹ Given that woman’s distinctive art is an art of appearing, it is perhaps unsurprising that character, which depends upon “not dissembling” (7:294), is unavailable to them. And given that women naturally focus on “pleasing others” (7:305), it is unsurprising that the self-governance required by the moral law is particularly out of character. Perhaps most striking is the subtle shift in terminology in Kant’s description of married life. In *Observations*, Kant suggests that “In marital life the united pair should [be] . . . animated and ruled by the understanding of the man and the taste of the wife” (2:242).¹¹⁰ In the context of an ethic that defines virtue as “the feeling of the beauty and dignity of human nature” (2:217), the cooperation of taste and judgment seems perfectly poised to constitute the “single moral person” (2:242) that Kant thinks a good marriage should involve; the woman’s taste can heighten the man’s feeling for human beauty, and the man’s understanding can heighten her feeling for human dignity. But in *Anthropology*, written almost 35 years later, Kant makes essentially the same point about mutual governance within marriage, but in terms with a strikingly different moral resonance: “Who, then, should have supreme command in the household? . . . the woman should *dominate* and the man should *govern*, for inclination dominates and understanding governs” (7:309). Not only has an emphasis on unity hardened into a description of mutual command, but more importantly, the woman’s role is no longer that of “taste” but that of “inclination.” While taste is part of virtue in Kant’s early ethics and a human characteristic that Kant continues to value throughout his life, “inclinations” in Kant’s mature moral philosophy lead one morally astray: “they are always burdensome to a rational being, and though he cannot lay them aside, they wrest from him the wish to be rid of them” (5:118).¹¹¹ Not only does Kant’s mature moral philosophy make it

¹⁰⁸ Pauline Kleingeld, in Kleingeld 1993, has pointed out that Kant denies that women have “courage” (7:303), which Kant claims is essential for virtue. This consideration seems less important to me than his points about character. I see no evidence that the courage that men have and women lack is the same courage needed for moral worth. Kant links “courage” with “physical power” and seems to have in mind pragmatic boldness.

¹⁰⁹ See Rumsey 1989 and Frierson 2006.

¹¹⁰ Kant adds, in striking contrast to *Anthropology*, that “In such a relationship a struggle for precedence is ridiculous” (2:242).

¹¹¹ To be fair, by the time that Kant wrote *Anthropology*, he had written his *Religion*, within which inclinations are subsumed under one or another predispositions to good. But Kant’s association of women with inclination dates back to his anthropology lectures of the 1770s and 1780s (see e.g. 25:718).

impossible for women to have a good will, but Kant even shifts in describing their non-moral advantages from the positively-connoting “taste” to the negatively-suggestive “inclination.”

In view of his apparent indifference to treating women as incapable of moral worth, it is perhaps unsurprising that when Kant turns to the details of his moral and especially political theory, he does not accord women rights equal to those of men. Within his political theory, women show up in two important contexts: citizenship and marriage. With respect to citizenship, Kant defines citizens as “the members of a society who are united for giving law” and insists on three essential

attributes of a citizen . . . : [1] lawful *freedom*, the attribute of obeying no other law than that to which he gives his consent; [2] civil *equality*, that of not recognizing among the *people* any superior with the moral capacity to bind him as a matter of right in a way that he could not in turn bind the other; and, [3] the attribute of civil *independence*, of owing his existence and preservation to his own rights and powers . . . (6:314)

Kant puts “all women” into the category of “passive citizens” or “mere associates” of the state, people who are not “fit to vote” but nonetheless have “freedom and equality *as human beings*” (6:314-5, cf. 8:295). And whereas Kant states as a general rule that the laws of a society must be such that “anyone can work his way of from this passive condition to an active [citizenship]” (6:315), his insistence that *all* women are passive implies that he excludes them from this condition. For Kant, women cannot vote and must remain forever dependent upon their husbands for representation in the public realm.

In *Anthropology*, Kant adds two important nuances to this account. First, he makes clear that while “woman regardless of age is declared to be immature in civil matters,” this is a specifically civil declaration; in fact, a wife is, if anything, “over-mature” in her ability to “represent both herself and her husband.” But “just as it does not belong to women to go to war, so women cannot personally defend their rights and pursue civil affairs by themselves.” Second, Kant suggests that women *do* in fact defend their rights and pursue civil affairs *indirectly*, since “this legal immaturity with respect to public transactions makes women all the more powerful in respect to domestic welfare; because here the *right of the weaker* enters in, which the male sex by its nature already feels called on to respect and defend” (7:209). Kant’s view seems to be that women are excluded from public politics not because of a genuine incapacity, but in order to empower them at home, where they can

control their husbands and thereby ensure that husbands take care of the family's public affairs.

Kant's political philosophy discusses marital rights in the general context of *property* rights, in a section illuminatingly entitled "on rights to persons akin to rights to things" (6:277). There are three ways to acquire "a person akin to a thing," when "a *man* acquires a *wife*; a *couple* acquires *children*; and a *family* acquires *servants*" (6:227). Of these three, Kant insists that both children and servants are acquired only for a specified period of time, after which they must be granted complete freedom from their parents/masters (6:281, 283). Only women are capable of being "acquired" for life.

To be fair, Kant's account of the husband's ownership right over the wife is carefully described not as a right to a person *as a thing*, which would blatantly contradict the obligation to respect all others as an end and not a mere means, but only as a right *akin* to rights to things. In fact, Kant's discussion of the marriage right is one of the clearest places where he articulated the view that women, despite whatever limitations they may have anthropologically, are nonetheless ends in themselves who have a "duty . . . to the humanity in [their] own person[s]" (6:280). It even follows from this, for Kant, that marriage rights – unlike rights over children and servants – must include entire reciprocity and "equality of possession" (6:278), such that just as the husband entirely owns the wife, so the wife in turn entirely owns the husband.

There is only one condition under which [it] is possible [to make oneself into a thing without compromising one's own humanity]: that while one person is acquired by the other . . . , the one who is acquired acquires the other in turn; for in this way each reclaims itself and restores its personality. (6:278).

Thereby Kant rules out polygamy, concubinage, prostitution, and "morganatic marriage" (whereby the husband withholds property or titles from the wife); and he even insists upon "equality in their possession of material goods" (6:278). Consistent with accounts of differences between sexes in *Anthropology* and elsewhere, Kant does not think that husband and wife play identical roles within marriage. In the *Metaphysics of Morals*, where legal rights are at stake, Kant insists that the husband "is to be [the wife's] master," which "cannot be regarded as conflicting with the natural equality of the couple if this dominance is based only on the natural superiority of the husband to his wife in his capacity to promote the common interest of the household" (6:279, see too 7:304). Consistent with his view that women hold power primarily through charming

manipulation of their husbands, Kant insists that the *law* recognize the husband as head of household, while the wife – through her relational adeptness – “dominates” the husband through his own will.

In the end, Kant’s account of the difference between the sexes is disturbing. Most blatantly, Kant was on the “wrong side” of the most important issues of the day, such as women’s education and citizenship. And while he “gallantly”¹¹² praises women’s distinctive charms, his overall account sees them primarily as means to the civil and moral development of *men*. Even Kant’s explicit endorsements of women’s equality (or superiority) fit within an overall attempt to defend and entrench patriarchal political structures. Saying that women have soft power that flourishes in contexts where they lack explicit and formal power is an excellent way to justify denying them political equality. And drawing attention to women’s ability to control men through charm is a good way of discounting the role of rational argument and dialogue at the level of intellectual equals. This discounting can have profound effects not only within marriages – where husbands will expect wives to be charming rather than wise – but also in the education of girls, which education would, for Kant, properly emphasize learning social graces rather than intellectual pursuits (including not only abstract metaphysics but also disciplines that involve more obvious uses of power, like physics, engineering, and politics).¹¹³

Simply accepting Kant’s views about women is unacceptable not only because they conflict with contemporary assumptions, but also because they conflict with Kant’s own transcendental anthropology.¹¹⁴ Through all the particular anti-feminist and misogynistic claims in Kant’s eventual account of women’s nature, it is the inability for women in their own right to have the unconditional worth of a good will that is the most morally and philosophically problematic. Moral responsibility, worthiness to be considered an end-in-itself, and the capacity for a good will are inextricably connected in Kant’s transcendental anthropology. Kant explicitly claims that women must be treated as ends-in-themselves (6:278, 280). And there is no evidence that he denies them moral *responsibility*. So Kant needs to provide an account of how

¹¹² A term that Kant constantly uses to describe his own accounts of women. See, e.g., 7:310.

¹¹³ One odd and striking feature of many of Kant’s descriptions of women’s education is the extent to which his accounts of the proper topics of learning for women match almost exactly his account of his own goals for pragmatic anthropology. In the actual practice of seeking to understand human beings pragmatically, Kant embodies his own highest ideals of *womanly* philosophy.

¹¹⁴ The rest of this discussion in this section focus on this general problem for his views. For some specific treatments of his accounts of women and marriage or women’s citizenship, see Denis 2007, Mendus 1992, and Shell 1996.

women can be capable of moral worth. And doing this will require substantial revisions in his anthropology.

* * *

Today, there are two major and opposing responses to Kant's characterization of women. The dominant response among those sympathetic to Kant "is to say that Kant's views on women are mistaken, that one should instead concentrate on his more important philosophical achievements, and that one can simply leave his theory about the sexes behind" (Kleingeld 1993:140). This response involves rejecting Kant's anthropological characterization of women and extending his descriptions of transcendental and even empirical anthropology to include women as well as men. Within the context of such an approach, one would affirm – with Kant – that character and rationality are crucial to virtue, but add – against Kant – that women are no less capable of these traits than men. A second response, dominant amongst feminist critics of thinkers like Kant, is to argue that Kant's philosophy as a whole reflects a masculinist bias. Such critics typically agree with Kant's general anthropological claim that there are important differences between men and women, but reject his identification of what is universal and normative for "human beings" with what is universal and normative for men. Carol Gilligan, for example, suggests – with Kant – that for women "morality is conceived in interpersonal terms and goodness is equated with . . . pleasing others" rather than in an "understanding of rights and rules" (Gilligan 1982:2).¹¹⁵ Like the Kant of *Observations*, however, Gilligan takes the "different voice" women bring to moral deliberation to be legitimate and needed to balance masculine emphasis on rule-following and personal autonomy (Gilligan 1982, see too Noddings 1984). While Gilligan focuses on differences in moral perspectives, other feminist thinkers have made similar points about Kant's transcendental anthropology more generally. His emphasis on reason and understanding over sensibility has been taken to reflect a

¹¹⁵ Importantly, Gilligan rejects the view that this difference is either universal or essential:

The different voice I describe is characterized not by gender but theme. Its association with women is an empirical observation . . . But this association is not absolute, and the contrasts between male and female voices are presented here to highlight a distinction . . . rather than to represent a generalization about either sex . . . No claims are made about the origins of the differences described or their distribution in the wider population, across cultures, or through time. Clearly, these differences arise in a social context where factors of social status and power combine with reproductive biology to shape the experiences of males and females and the relations between the sexes. (p. 2)

(For a more essentialist account of gender differences that shares many claims in common with both Gilligan and Kant (though situated in a psychoanalytic framework), see Chodorow 1974.)

masculinist bias in epistemology, one that puts “the Enlightenment conception of a universal, rational subject” above “feminist notions that the self is embedded in social relations, that the self is embodied, and is thus historically specific and partial” (Schott 1997:8).

This divided response to Kant’s thought highlights in contemporary form the problem that arises within Kant’s own anthropology. Philosophers who find Kant’s transcendental anthropology convincing have sought to jettison the empirical-anthropological accounts of women than make them seem ill-suited for fulfilling the requirements of that transcendental anthropology. But philosophers and empirical psychologists who study sex and gender often end up supporting, if not Kant’s specific claims, at least accounts of gender differences that raise similar philosophical problems for a broadly Kantian account of moral and epistemic norms. What seemed to be a tension internal to Kant, one that he lamely resolved by simply settling into misogyny and ignoring the problems this raised for his transcendental anthropology, appears as a real problem for anyone who finds plausible both Kant’s arguments for universal norms governing thought, choice, and feeling and the importance of empirical sensitivity to human differences in thinking about how those norms play out in the real human world.¹¹⁶

One important way to deal with the tension – for the present day, even if not to salvage Kant’s own thinking – would be through more careful and fine-grained approaches to both transcendental and empirical anthropology. These might find that women and men are not different in ways that have moral relevance. Even if, say, morality requires acting on principles and women tend to be more situational, this does not *necessarily* mean that they lack what is necessary for morality. Morality might *also* require situational sensitivity,¹¹⁷ and women might

¹¹⁶ Incidentally, this problem is more serious, and less easily resolved, than the problem that Gilligan emphasizes in her critique of Kohlberg. Gilligan is a psychologist writing in response to Kohlberg’s (and others’) theories of human development, which used empirical studies of human beings from childhood to adulthood to develop scales for measuring such things as moral and emotional maturity. In a context where the basis for the dominant normative claims about what human beings ought to be is largely driven by a description of how “eighty-four *boys* . . . develop[ed] . . . over twenty years” (18), Gilligan’s observations that girls and women develop differently is a serious ground for calling into question the normative claims themselves. But Kant does not defend his normative claims on the basis of observing the actual development of human beings. Instead, he defends them from-within, as conditions of the possibility of any legitimate claims to knowledge or moral responsibility or aesthetic experience. Thus empirical evidence showing that one sex fails to meet these standards provides no ground for rejecting them.

¹¹⁷ Barbara Herman has recently argued that Kant’s moral theory depends at least as much upon a situational sensitivity that is highly emotional as it does upon the categorical imperative itself. See Herman 1993.

also be capable of morally principled action, even if not in the same way or to the same degree as men.

Another middle-ground would accept both Kant's philosophical defense of norms that seem more masculine than feminine and empirical evidence of differences between sexes, but reject the teleological essentialism underlying Kant's explanation of those differences.¹¹⁸ As Carol Gilligan emphasizes, "differences arise in a social context where factors of social status and power combine with reproductive biology to shape the experiences of males and females and the relations between the sexes" (Gilligan 1984:2). Kant himself recognizes that the sex differences he discusses arise only in particular social and political contexts (7:304) and his discussion of women's civil inequality (6:314, 7:204) remains at least open to the idea that women's civil immaturity is socially-created rather than natural. Differences between men and women that inhibit women from fully realizing Kantian ideals of autonomy may be due to unjust social conditions that can and should be remedied. This approach could give rise to an increased attention to the social and political reforms that could create a world within which women would have as good a chance as men at measuring up to universally human moral (and other) norms. Kant, unfortunately, rejected this middle-ground in both his anthropology – where he insists that differences between men and women are natural and not merely social (e.g. 25:709) – and in his politics – where his explanation of the passive citizenship of "all women" is combined with both complacency and an account of marriage that seems to reinforce this civil inequality.

Finally, even if it does turn out that there are essential differences between men and women *and* that these differences make it considerably more difficult for women to attain to a good will, one might still – and Kant certainly should – insist that it *is* possible for women to have unconditional moral worth. Modifying Kant's claim from *Religion*,¹¹⁹ we might say "In spite of [one's sex], the command that we ought to become better human beings still resounds unabated in our souls; consequently, we must also be capable of it" (6:45). In this context, the difference between men and women would be akin to differences between temperaments, not a denial of the *possibility* of virtue for women, but a detailed attention to the fact that women will face greater and different

¹¹⁸ The recent emphasis on distinguishing "sex" and "gender" obviously fits into this general approach. Insofar as one's gender can be distinguished from one's biological sex, one need not identify biological sex (male/female) with the characteristics associated with particular genders (masculine/feminine), and one can largely ascribe gender characteristics to social factors.

¹¹⁹ See chapter three, p. xxx.

challenges in their progress towards virtue than men. One might even recast the differences between the sexes in terms of different manifestations of radical evil. Whereas men's radical evil typically manifests itself in an unsocial sociability rooted in overwhelming love of honor that gives rise to passions of jealousy and rivalry, women's radical evil might more typically manifest itself in a desire to please that encourages duplicity and sacrificing principle for the sake of relationship. The claims about women that seem to preclude her from having true virtue are certainly no more extreme than Kant's claims about radical evil in human nature in general; Kant just lacks an account of the emergence of women from their particular weaknesses. Again, however, Kant did not seem to have seen the difference between the sexes in these terms. Despite the change in his moral theory that made "feminine virtues" mere illusions of virtue, Kant never gave up the idea that women ought to seek these virtues rather than the unconditional good will to which men should aspire.

In the end, Kant's treatment of women, in the context of his anthropology as a whole, raises problems and tensions that continue even today to affect thinking about relations between the sexes, and between empirical and transcendental philosophy. But Kant not only rejects the most natural ways of dealing with these problems, but his infatuation with the "charming difference that nature sought to establish between the two human sexes" (2:228) seems to have made him blindly and complacently unaware of them.

IV. Racial and ethnic differences

Kant's account of sex differences is not the only part of his anthropology that is offensive and problematic for his philosophy. Another important aspect of Kant's account of human beings is his theory of racial and national character. It is important to distinguish three separate aspects of Kant's account, especially with respect to racial difference. First, Kant makes statements about other races that are, from our contemporary standpoint, outrageously racist. Second, Kant develops a complicated *theory* of race, one that played a role in the development of scientific racism in the 19th century and thereby continues to affect the way races are conceived today. Finally, Kant's moral and political theory at times specifically addresses the relationships between peoples of different racial and ethnic groups. Precisely how these different elements of his views fit together is not always clear; and at least the first, and probably the second, share with

Kant's views on women both immediate offensiveness and serious tension with his transcendental philosophy.

Throughout this section, I focus on racial rather than ethnic (or what Kant calls "national") differences.¹²⁰ Kant was (one of) the first thinker(s) to develop a scientific concept of "race," and many of his most outrageous comments about other peoples are comments about non-European races. But it is important to note that with the exception of a series of articles published during the 1780s, Kant was much more interested in differences amongst European peoples than in differences between Europeans and non-Europeans. His published *Anthropology* includes a major section on differences amongst European nations but only two short paragraphs on the character of the races, and even his *Observations on the Feeling of the Beautiful and Sublime* devotes only a "quick look" at "other parts of the world" (2:252) after offering a substantial discussion of differences between European people groups (French, Spanish, English, etc.) (2:243-52). The emphasis of this section on different races is due both to the presence of a systematic race-theory in Kant's own writings and to the importance of "race" today. For Kant, however, differences amongst Europeans were at least as important as differences between Europeans and others.¹²¹

a) Kant's descriptions of other races

Kant's most disgraceful (published) claims about races are found in his *Observations on the Feeling of the Beautiful and Sublime*. After describing a conversation about the nature of women between an African carpenter and a European missionary, Kant writes of the African's comments,

There might be something here worth considering, except for the fact that this scoundrel was completely black from head to foot, a distinct proof that what he said was stupid. (2:254-5)

¹²⁰ I also do not discuss Kant's important comments about Jews (see especially R 6:125ff.); some, but not all, of the present discussion is relevant to assessing these comments. While his views have been seen as important precursors of later German anti-Semitism, they would take the present discussion too far afield because Kant sees the Jews primarily through a cultural and religious rather than racial lens.

¹²¹ Whether Kant would extend this interest in ethnic as opposed to racial differences to other races is not clear. He often distinguishes different ethnic groups within non-European races (see, e.g., 2:252), but he also seems to think that a "definite national character" is an accomplishment that depends upon a level of civilization of which non-European races seem incapable (see, e.g., 7:319).

Kant's general characterization of Black Africans in *Observations*, though painfully offensive, is also worth quoting:

The **Negroes** of Africa have by nature no feeling that rises above the ridiculous. Mr. **Hume** challenges anyone to adduce a single example where a Negro has demonstrated talents, and asserts that among the hundreds of thousands of blacks who have been transported elsewhere from their countries, although very many of them have been set free, nevertheless not a single one has ever be found who has accomplished something great in art or science or shown any other praiseworthy quality, while among the whites there are always those who rise up from the lowest rabble and through extraordinary gifts earn respect in the world. So essential is the difference between these two human kinds, and it seems to be just as great with regard to the capacities of mind as it is with respect to color. The religion of fetishes which is widespread among them is perhaps a sort of idolatry, which sinks so deeply into the ridiculous as ever seems to be possible for human nature. A bird's feather, a cow's horn, a shell, or any other common thing, as soon as it is consecrated with some words, is an object of veneration and of invocation in swearing oaths. The blacks are very vain, but in the Negro's way, and so talkative that they must be driven apart from each other by blows. (2:253)

This text does not represent the limit of Kant's offensive comments. In notes from his lecture course on *Anthropology*, Kant claims, in terms reminiscent of his comments about women:

If we compare the character of the Oriental nations with the character of the Europeans, we here thus find an essential difference, which among all the governments and variations has nevertheless remained in the case of the Oriental nations. A capacity to act in accordance with concepts and principles is required for character. All Oriental nations are completely incapable of judgment in accordance with concepts. It is a big difference to judge a matter according to shape, appearance, and intuition, and to judge [it] according to concepts. All Oriental nations are not in the position to explain a single property of morality or of justice through concepts; rather all their morals are based on appearance. (25:655)

And in his *Physical Geography*,¹²² Kant offers a sort of summary of his views of the different races of the world:

¹²² Kant taught a course in Physical Geography for decades (more than any other single course), and, like his lectures on anthropology, student transcripts of these lectures were circulated during Kant's days. Until

Humanity is at its greatest perfection in the race of the whites. The yellow Indians do have a meager talent. The Negroes are far below them and at the lowest point are a part of the American peoples. (*Physical Geography*, 9:316, see too 25:843,¹²³ 1187-8)

These comments offer but a sample of the dismissive and demeaning views about non-Europeans scattered throughout Kant's writings and lectures. Now Kant is not always *entirely* dismissive of other races. In *Observations*, in sharp contrast to his *Physical Geography*, Kant has some admiration for the "sublime cast of mind" of the "**savages** . . . of **North America**."

Lycurgus probably gave laws to such savages, and if a law-giver were to arise among the six [Native American] nations, one would see a Spartan republic arise in the new world; just as the undertaking of the Argonauts is little different from the military expeditions of these Indians, and **Jason** has nothing over **Attakakullakulla** except the honor of a Greek name. (2:253-4, though cf. 25:1187)

And elsewhere Kant says that "the Hindus . . . have a strong degree of composure, . . . they all look like philosophers, . . . [and] they acquire culture in the highest degree" (25:1187). On the whole, however, Kant's informal "observations" about other races reflect the prejudices of an European satisfied with the superiority of his own race and ready to believe the worst and most degrading claims about other races.

* * *

To those who know Kant through his moral philosophy or the universal claims of his transcendental and empirical anthropology, these deeply disdainful comments about other races are disturbing to say the least. How should we respond to comments that seem so out of line with the respect for humanity that Kant emphasizes elsewhere in his work?

recently, the main form in which these lectures have been available to Kant scholars has been in an edition put together by Kant's student Friedrich Theodor Rink, likely based on notes from two different courses and included in volume 9 of Kant's works. Volume 26 of Kant's works (published partly in 2009 and partly forthcoming) will include many more versions of student notes from Kant's Physical Geography course.

¹²³ The remarks at 25:843, while not wholly reliable because based on lecture notes, are perhaps the most disturbing. There Kant not only sets up an implicit hierarchy of races, but adds that "Negros are not capable of any further civilization . . . The Indians and Chinese seem to be static in perfection, for their history books show that they do not know more now than they have long known." (25:843). For both Africans and Asians, Kant alludes to character traits that are not merely physical (civilization and learning) and suggests not only that these races are inferior to whites, but that they cannot ever improve.

Unlike Kant's comments about sex, where many continue to argue that there are important differences between men and women that may be relevant to moral or epistemic issues, the notion that there are serious innate differences between races that would inhibit members of a particular race from being able to satisfy the demands of Kant's transcendental anthropology can hardly be taken seriously. To those who know Toni Morrison, Jacob Lawrence, Benjamin Banneker, or Wangari Maathai, Kant's reference to Hume's claim that "not a single [African] has ever be found who has accomplished something great in art or science" would display little more than laughable ignorance if it were not so appalling (2:253). With respect race, the issue is not whether to accept Kant's racial distinctions and adjust his transcendental philosophy or vice versa. No serious thinker today can affirm Kant's racial observations. But two issues remain: first, whether Kant's views on non-white races taint the rest of his philosophy such that his claims about, say, moral norms or cultural progress must be abandoned because they are inextricably linked with racism, and second, how Kant – that champion of universal human dignity – could espouse views that seem to deny that dignity to most of the world.

One response to the first issue involves simply dismissing or ignoring Kant's racially offensive comments. While there are glimmers of Kant's views of other races in his more well-known writings – the most famous being his reference in *Groundwork* to "the South Sea Islanders . . . [who] let talents rust and are concerned with devoting life merely to idleness, amusement, [and] procreation" (4:423) – by and large Kant's best read texts give little explicit indication of his racial views. It seems easy to excise these offensive texts from Kant's corpus, ignore them, and focus on the parts of Kant's thought that can and should be candidates for serious consideration today. This strategy is far and away the dominant way of dealing with Kant's claims about races and has important advantages, but it is not without its dangers.

The advantages should be clear. Kant's philosophy has had profound impacts on metaphysics, epistemology, ethics, aesthetics, philosophy of religion, political theory and so on. Kant's moral philosophy, in particular, continues to play important roles in safeguarding human rights and individual autonomy. Throwing out Kant's insights in these areas because of his personal views about other races is a waste. Moreover, Kant himself provides an important justification for leaving his racially offensive comments behind and focusing on the rest of his anthropology. For Kant, transcendental anthropology (epistemology, ethics, and even aesthetics) must be developed a priori; empirical insights – including observations about

human differences – are relevant only later, in thinking about how to apply a priori norms to empirically-situated human beings. Given Kant’s own discipline in isolating his transcendental philosophy from his empirical observations, one seems justified in ignoring the latter and reaping the insights of the former, especially with empirical claims so clearly false.

That said, simply ignoring Kant’s comments on race brings important dangers. One danger is that one risks misunderstanding those parts of Kant that one chooses to accept. Kant’s apparently off-handed reference to the South Sea Islanders, for example, actually involves him taking an important stand in contemporary debates about the moral status of so-called “primitive” peoples. Most travelers’ accounts, especially of Tahiti (rediscovered by Europeans in 1767), were both “morally provocative” – contrasting Europeans morals with those of the native – and “complimentary,” in that travelers to these places typically presented the lives of at least the Tahitians (and often other “savages” as well) as idyllic not only in terms of pleasures but also of morals (Wilson 1998: 317). Kant’s insistence that the cultivation of one’s perfections (including non-moral talents) is a duty is an important part of his moral philosophy, one for which Kant offers an apparently universal justification. But in the context of Kant’s claims about other races, this claim can be situated into a general Kantian defense of the superiority of the ambitiously progressive historical self-conception of a European comparing himself to the rest of the world. In itself, this added insight does not give a reason to *reject* Kant’s claim that humans ought to cultivate their talents. But it does force one to look more carefully at the justification and implications of that claim.¹²⁴ Precisely how to understand claims that might otherwise seem to be universal, especially when these arise in Kant’s transcendental anthropology, cannot be easily settled simply by pointing to Kant’s views about races. But paying attention to these views can force more careful attention to nuances in Kant’s philosophical views.

This point highlights other risks of simply dismissing Kant’s views on race. Kant’s claims about other races at least *seem* to conflict with

¹²⁴ For another example, when Kant claims in the *Critique of Judgment* that human beings seek systematic interconnections amongst empirical objects and the laws that govern them, he might seem to mean this as a universal claim about all human beings. But Kant’s views about other races (e.g. about “Oriental nations”) suggests that it might simply be a claim about the human species as a whole. Kant may not be committed to the view that each human being seeks systematic interconnection, but only that human beings as a group seek this, and within that group, that white Europeans lead the way.

other aspects of Kant's anthropology. At the very least, they raise questions about how Kant could have reconciled his universal anthropology with a view that "Negroes" are irredeemably stupid and "Orientals" incapable of concepts. Trying to figure out how Kant could have held together what seem to be such disparate views can open up new insights into the meaning, limits, and dangers of what might otherwise seem benign aspects of his philosophy.¹²⁵ Alternatively, showing precisely the way in which, say, Kant's moral theory conflicts with these claims can reveal that moral theory as an important resource for overcoming racism today.¹²⁶ Finally, failing to pay attention to Kant's negative views on race can mark a missed opportunity to more fully understand the limits of philosophy itself. Investigating how Kant, who claims, "I would feel by far less useful than the common laborer if I did not believe that [my philosophy] could impart a value to all others in order to establish the rights of humanity" (20:44), could hold such offensive views about other races can help reveal some of the causes that continue to prolong racism today.

Despite the dangers of simply *ignoring* Kant's racially offensive comments, however, dismissing Kant's whole philosophy as tainted by racism is even more dangerous. For one thing, Kant's transcendental anthropology has played an extremely important role in helping philosophers – and our society as a whole – come to see racism as unacceptable. Kant's moral philosophy, with its emphasis on the need to respect the infinite worth of each and every human being, is still one of the most powerful philosophical tools for combating racism. And as we will see in chapter nine, even Kant's epistemological claim that our world is constructed in terms of a priori categories we impose on it has helped cultivate an awareness of the extent to which *nonuniversal* categories of thought can also shape our experiences, an awareness crucial for cross-cultural understanding. And although Kant's interest in transforming the study of other cultures into a serious academic discipline was tainted by how he wanted to do that, his conviction that being an educated world citizen requires understanding not only universal characteristics of human beings but also human differences remains an important insight today. Finally, a considerable amount of Kant's philosophy, at least as that philosophy has been taken up and influences philosophers today, can be freed of Kant's racist views. The mere fact that most readers of Kant – including those whose interpretations are most influential – are virtually (and often completely) unaware of his views on races shows that those views can, at least to a considerable degree, be understood and

¹²⁵ See Eze 1994, Larrimore 1999, and Loudon 2000.

¹²⁶ See Boxill and Hill 2001, Loudon 2000.

appreciated independently. As philosophers and scholars continue to explore the implications and impact of Kant's racial views on his philosophy as a whole, some interpretations of Kant will have to change, and some aspects of his views that might have seemed plausible will now raise more suspicion. But there is at present no reason to think that his philosophy as a whole will need to be dismissed simply because most of his claims about other races must be.

When we turn to how Kant could have held these views, the obvious response – simply ascribing Kant's racial stereotypes to his eighteenth-century background – is not entirely satisfying. On the one hand, his racist views are largely informed by popular prejudices of Kant's time. The contexts in which Kant's comments occur are almost entirely contexts in which Kant is deliberately seeking to write (or teach) in ways that will be “popular,” “entertaining” for his readers and students (10:146). *Observations*, in which his most outrageous claims occur, was Kant's most popular book during his lifetime and was written in part to attract students to his lectures. Because negative attitudes towards non-white racial groups were widely shared amongst the public at large in the eighteenth century, Kant's demeaning comments would likely have enhanced his works' popularity. His working-class background may have encouraged him to draw divisions between people that would put himself and the upper class students that he needed to attract to his lectures on the same side. Being a relatively poor intellectual who never went more than 90 miles from home, Kant was limited in his data about other races to accounts written by merchants, explorers, and missionaries. In fact, Kant's courses in *Physical Geography* and *Anthropology* were, at least in part, designed “to make a more certain knowledge of believable travel accounts and to make this into a legitimate course of study.”¹²⁷ Thus Kant was, to a considerable degree, limited by the biases and prejudices of the travel accounts to which he had access and the culture of which he was a part.

On the other hand, however, it is also clear that eighteenth-century thinking about non-white people was not *uniformly* negative. Kant's empirical sources were often much more generous in their observations than Kant, being written by travelers influenced by a moral ideal of simplicity that seemed well-exhibited in the exotic peoples they observed (Wilson 1998). Among the most important alternative theoretical ways of thinking about races were those of Kant's own student Herder, who, drawing from similar travel logs and empirical sources, developed a much less patronizing view of non-European

¹²⁷ Wilson 2006: 3.

nations, and Georg Forster, who not only published travel accounts of his own that emphasized much more positive views of non-Europeans but also specifically criticized Kant's own race theory as being insufficiently egalitarian in 1786. In both cases, Kant fought *against* more generous portrayals of other races. In his review of Herder's *Ideas*, Kant even wrote:

[F]rom a multiplicity of descriptions of countries one can prove, if one wants to, that Americans, Tibetans, and other genuine Mongolian peoples have no beard, but also, if it suits you better, that all of them are by nature bearded . . . ; that Americans and Negroes are each a race, sunk beneath the remaining of the human species in their mental predispositions, but on the other side by just as apparent records that as regards their natural predispositions, they are to be estimated equal to every other inhabitant of the world; so it remains to the choice of the philosopher whether he wants to assume differences in nature or wants to just everything in accordance with the principle "Everything is as it is with us." (8:62)

This self-awareness about the process of picking and choosing amidst empirical data shows how Kant *could* – even with his sources and his cultural baggage – have developed a different view of non-white races. And Kant should have seen the inconsistency of his dismissal of other races with his own personal and philosophical trajectory: growing up in a working class family, rising into the ranks of the intelligentsia, and then recognizing on reading Rousseau that intellectuals are “far less useful than the common laborer” except insofar as they defend “the rights of humanity” (20:44). In the end, while one can point to reasons for Kant's demeaning views of other races in his cultural context, Kant's own words require him to acknowledge “the choice of the philosopher.” And while one might understand this choice in terms of pleasing the crowd (especially given the public purpose of his *Observations*)¹²⁸ or even in terms simply of making his best attempt at getting things right (given the

¹²⁸ One might also see Kant not as endorsing but merely exploiting the racially demeaning views of his times. Kant's lower class background may have provided an incentive to embrace *these* demeaning views to deflect class-based prejudices of his contemporaries. Moreover, by portraying Europeans as superior by virtue of their ability to think in terms of concepts, to govern their lives on the basis of principles, and so on, Kant exploits his readers' (and listeners') prejudices to exhort them to personal moral reform. Not only are the norms of Kant's transcendental anthropology the routes to autonomy and the fulfillment of one's vocation as a human being, but they are also the ways to distinguish oneself as a European from the savage Americans, stupid Negroes, and fantastical Orientals. If one assumes that one's readers and listeners already have a sense of white superiority, associating that superiority with the authentic norms of human autonomy is a way to inspire his readers to pursue these norms. In that sense, Kant masterfully uses the rhetorical strategy whereby his readers mock and deride attributes of others, only to find that they need to overcome those same attributes in themselves.

pedagogical purpose of his physical geography and anthropology lectures), one must also admit that Kant, despite his acuity in some areas of philosophy, was neither sufficiently thoughtful nor sufficiently courageous in thinking about other races. Whatever the explanation, Kant's ability to combine a transcendental anthropology that emphasizes universality with utterly dismissive claims about most of humanity is an important lesson in the limits of merely philosophical insight.

b) Kant's Race Theory

Kant's most extreme claims about race come in early or informal works, largely disconnected from any formal theory about races. But over a decade after the publication of *Observations*, Kant wrote an essay entitled "Of the difference races of human beings" that marked the beginning of a series of papers in which Kant "invented the concept of race" (Bernasconi 2001:11) or at least became "a leading proponent of the concept of race at a time when its scientific status was still far from secure" (Bernasconi 2002: 146). Kant's racially offensive but informal remarks bear a greater superficial similarity to present-day racism. But his less immediately offensive *theory* of race arguably played a more significant role in actually creating the conditions for present-day racism by giving "the concept [of race] sufficient definition for subsequent users to believe that they were addressing something whose scientific status could at least be debated" (Bernasconi 2001: 11). As one commentator puts it, "Once Kant's role in constructing a rigorous concept of race is established, it is relatively easy to give Kant a place in the history of racism" (Bernasconi 2002: 146), a place he would not warrant merely for his offensive comments (comments that played little role in the development of racism and that were common at the time).

In one important respect, Kant's theory of race was deeply *anti-racist*, in that Kant was a staunch defender of "monogenesis" – the view that all human beings are a single species derived from a common ancestor – during a period in which polygenesis – the view that, for instance, black Africans and white Europeans are actually different *species* – was gaining prominence as Europeans increasingly interacted with different groups. The different behavior and physical appearance of distant peoples challenged the limits of the European imagination to the point that postulating that different peoples were different species seemed a natural response to the discomfiting possibility that "we" and "they" were the same sort of being. At the same time, the basic categories of biological science were in flux, so there was no universally accepted

criterion for determining commonality of species. Neither “common descent” nor taxonomic similarity provided clear bases for determining whether different human populations were descended from common ancestors and/or sufficiently similar to be grouped together. The heritability of racial characteristics even after members of different members were transplanted to new climates – i.e., the fact the Europeans and their children did not become dark-skinned when living in Africa – even seemed to suggest polygenesis.

In response to the growing interest in polygenesis, Kant sought a scientific account of the human species that would reconcile monogenesis with European’s desire to distinguish between people with recognizable and heritable differences.¹²⁹ The essence of Kant’s account is to distinguish concepts of “species” and “race” and to provide clear criteria for each. With respect to “species,” Kant adopts from his contemporary Buffon what has come to be the standard account in contemporary biology¹³⁰: “animals that produce fertile young with one another (whatever difference in shape there may be) still belong to one and the same physical species” (2:429, see too 8:165-9). Kant immediately applies this to the human case: “According to this concept, all human beings on the wide earth belong to one and the same natural species because they consistently beget fertile children with one another” (2:430). Because being the same species does not *necessarily* imply common ancestry, Kant makes his affirmation of (and justification for) monogenesis explicit: “[H]uman beings belong not merely to one and the same *species*, but also to one *family*, [since otherwise] many local creations [of members of the same species] would have to be assumed – an opinion which needlessly multiplies the number of causes” (2:430).

¹²⁹ It is not entirely clear why Kant was motivated to defend monogenesis. He may have had religious motivations, a desire to ensure that biology is reconcilable with the Biblical account of creation, but given Kant’s willingness in other areas to interpret the Bible to fit “within the boundaries of mere reason,” this reason seems unlikely. Another theoretical possibility would be a resistance to the prejudice and oppression that were often justified on the basis of polygenesis. But although (as we will see below) Kant was vehemently opposed to the dominant forms of explicit racial oppression in his day, this justification seems unlikely given the racial views discussed in the previous subsection. Most likely, Kant’s reason for defending monogenesis is just what he claimed it to be, that polygenesis unnecessarily multiplies scientific hypothesis (see 2:430). That is, Kant seems to have had a basically scientific motivation for developing his view of different races. At least one other plausible and likely cause, suggested by Bernasconi, is that this problem in biology – how to reconcile variation with monogenesis – simply gave Kant a good opportunity to show the power of his philosophy of biology to deliver concrete results to a real problem facing the scientific world.

¹³⁰ There is considerable debate about and nuance within this present-day biological species concept, but the concept developed by Buffon and Kant provides the core of the most prevalent contemporary versions of it.

Having settled the issue in favor of monogenesis, Kant needs to explain heritable diversity between different human groups, and he does so with his concept of a race, defined as follows:

Among . . . the hereditary differences of animals which belong to a single [species], those which persistently preserve themselves in all transplantings (transpositions to other regions) over prolonged generations among themselves and which also always beget half-breed young in the mixing with other variations of the same [species] are called *races*. (2:430)

In order to distinguish racial characteristics from differences due merely to environmental conditions, Kant insists that racial differences must persist over many generations even after the relevant “race” is transplanted to a different place. Racial differences must also *blend* when members of different races interbreed, which allows Kant to distinguish such differences from what he calls difference in “strain” (2:430) and thereby avoid calling different white European peoples different “races.”

In this way, *Negroes* and *whites*, while not different kinds of human beings . . ., are still two *different races* because each of the two perpetuates itself in all regions and both necessarily beget . . . *blends* (mulattoes) with one another. By contrast, *blondes* and *brunettes* are not difference *races* of whites, because a blond man can have entirely blond children with a brunette woman. (2:431)

As this example suggests, Kant goes on to argue that *skin color* “is especially suited” for dividing the races, since “no other characteristic property is *necessarily hereditary* . . . [and] in the mixing of [peoples with different skin colors] the character of each one is *unfailingly hereditary*” (8:94-5). Thus Kant divides the human species into “four classificatory differences . . . with respect to skin color . . . [:] the *whites*, the *yellow* Indians, the *Negroes*, and the *copper-red* Americans” (8:93).¹³¹

For Kant, however, merely defining the concept of race and classifying human beings is insufficient. Kant also aims to show how the concept of race can be illuminated by his overall philosophy of biology, and in particular by the role of natural teleology in empirical anthropology. Thus Kant uses his account of natural predispositions in

¹³¹ Strikingly, Kant says – within a single page – both that “how much the natural skin color of the Kaffirs differs from that of the Negroes . . . will not be decisively settled for a long time” and that “I assume that there are no more hereditary ethnic characters . . . than the above four” (8:93-4). Over the course of his writings on race, Kant never changes his definition of race, focus on skin color, or insistence that there are only four races; but he changes his account of *which* four races are irreducible (cf., e.g. 2:432, 8:93).

order to explain both *how* and *why* human beings became differentiated into different races. Starting with the *purpose* of racial differentiation, Kant claims, “The human species was destined for all climates and for every soil; consequently, various germs and natural predispositions had to lie ready in him to be on occasion either unfolded or restrained” (2:435, see too 8:168). The idea is that different racial characteristics are well-suited for different climates. Given Nature’s end – for humans to settle the entire globe – she endowed human beings with a variety of predispositions that could develop differently in accordance with different local conditions “so that he would become suited to his place” (2:435). For Kant, however, variability with local conditions is importantly not *merely* an ability to adapt; it has a hereditary component:

Once a race . . . had established itself . . . this race could not be transformed into another one through any influences of the climate. For only the [original representatives of the species] can [develop]¹³² into a race; however, once a race has taken root and suffocated the other germs, it resists all transformation just because the character of the race has then become prevailing in the generative power. (2:442, see too 8:166, 172)

What begins as a mere lack of expression of certain natural “germs” (akin to predispositions) becomes, over time, a “suffocation” of those germs. It is unclear whether Kant intends to say that the germs literally die out, or – more likely – that a propensity for them not to express themselves becomes hereditary. In either case, individual adaptations to climate become fixed characteristics. And this, too, has a natural purpose, so that human beings, “over the course of generations . . . appear to be . . . made for that place” in which they reside (2:435). Nature intends not only for human beings to spread all over the globe, but also for humans to fit well wherever they find themselves to be. Kant thus charts a middle path between those who claim that human beings are biologically distinct and those that claim that differences are environmental. Differences between human beings are caused by environmental factors, but at least some of these differences – notably skin color – can become hereditary.

In many respects, Kant’s formal theory of race is much less problematic than his informal negative comments about various races.

¹³² Kant uses the technical term “degenerate” here. In 18th century biology, the term “degeneration” was used in virtually the way that we use the term “evolve” today, and it lacked many of the negative connotations that we associate with the term today.

With rare exceptions, Kant's race essays refrain from describing moral or intellectual qualities as hereditary, and the claim that skin color is necessarily hereditary is not, in itself, particularly offensive. Arguably, Kant's account of race is even an important step towards a broadly Darwinian account of the possibility of environmentally-caused heritable changes in given populations. But Kant's race theory raises three new problems for assessing Kant's philosophy as a whole. First, given how Kant situates his race theory in the context of the stagnation of various natural predispositions, his race theory raises the stakes of his racially disparaging comments. If other races have literally lost the *capacity* for moral or intellectual advancement, this poses problems for Kant's moral theory and philosophy of history that are similar to those raised in the context of the sexes (where women seemed incapable of moral worth). Second, whatever its relationship to his moral theory, Kant's race theory seems deeply intertwined with the account of natural teleology in his *Critique of Judgment*, which provides the capstone of his transcendental anthropology as a whole. Even if Kant's transcendental anthropology could be isolated from Kant's informal comments about races, it seems harder to isolate his race *theory*. Finally, even if Kant's race theory is not as immediately offensive as the comments discussed in the last section, by contributing to the development of a scientific, skin-color-based conception of race, Kant arguably played a real historical role in the development of modern racism.

With respect to the first issue, Kant's works provide mixed evidence about the extent to which he conceived of moral and intellectual attributes as irremediably fixed in races. In *Observations*, the early essay with Kant's most atrocious claims about races, Kant adds a crucial footnote to the title of the section in which he discusses different races¹³³:

My intention is not at all to portray the characters of the peoples in detail; rather I will only outline some features that express the feeling of the sublime and the beautiful in them. One can readily guess that only a tolerable level of accuracy can be demanded in such a depiction, that its prototypes stand out in the large crowds of those who make claim to a finer feeling, and that no nation is lacking in casts of mind which unite the foremost predominant qualities of this kind. For this reason the criticism that might occasionally be cast on

¹³³ Because this section focuses on "The Character of the *Nations*," including non-European races as an afterthought, the footnotes *might* be restricted to Europeans. But they show a way that Kant could reconcile claims about races with his universalist moral theory.

a people can offend no one, as it is like a ball that one can always hit to his neighbor. (2:243n)

Later, Kant reiterates,

It is hardly necessary for me to repeat my previous apology here. In each people the finest portion contains praiseworthy characters of all sorts, and whoever is affected by one or another criticism will, if he is fine enough, understand it to his advantage, which lies in leaving everyone else to his fate but making an exception of himself. (2:245n)

At least in *Observations*, Kant sees differences between people groups like those between temperaments, as natural advantages or disadvantages that can be overcome through making an exception of oneself. Readers should take negative characterizations of their nation or race as exhortations to moral strength rather than signs of inextricable inferiority. Moreover, although Kant refers to differences between Europeans and Africans as “essential” (2:253) and seems fixed even upon change of conditions, his early comments on racial differences stick to the level of “observations.” Kant even claims,

I will not investigate here whether these national differences are contingent and depend upon the times and the type of government, or whether they are connected with a certain necessity with the climate. (2:243n)

And even when Kant develops his formal race theory, in which he argues for essential and hereditary racial characteristics, he insists that “no characteristic property other [than skin color] is *necessarily hereditary*” (8: 94). And in an anthropology lecture delivered the same year that Kant published his “Of the different races of human beings,” he uses Indians as examples to show that *all* human beings have the same underlying “germs.”

Who has seen a savage Indian or Greenlander, should he indeed believe that there is a germ innate to this same [being], to become just such a man in accordance with Parisian fashion, as another [would become]? He has, however, the same germs as a civilized human being, only they are not yet developed. (25:694)

In these and other passages, Kant seems to endorse the view that whatever moral and intellectual differences there are between races can, to some extent, be overcome.

On the other hand, Kant's comments about other races often imply a heritable and unchangeable moral and intellectual inferiority of some races to others. Kant's rankings of various races (see, e.g., 2:441), and his claim in *Observations* that the differences between Africans and Europeans are "essential" and "just as great with regard to the capacities of mind as it is with respect to color" (2:253) imply as much. And although his formal race essays emphasize skin color, the last of these essays describes Native Americans as "*incapable* of any culture" (8:176, emphasis added, cf. 10:239), and the first ascribes to them a "half-extinguishes life power" while describing Africans as "lazy, soft, and trifling" (2:438). Perhaps the most systematic-sounding claim comes in a footnote of Kant's last race essay. Here, in the context of a discussion of whether African slaves could be used as free laborers, Kant writes:

Should one not conclude . . . that in addition to the *faculty* to work, there is also an immediate drive to activity (especially to the sustained activity that one calls industry), which . . . is especially interwoven with certain natural predispositions; and that Indians as well as Negroes do not bring any more of this impetus into other climates and pass it on to their offspring than was needed for their preservation in their old motherland . . . [where t]he far lesser needs . . . demand no greater predispositions to activity. (8:174n)

Strictly speaking, Kant does not claim that Negroes or Indians are *incapable* of activity or industry, but he does suggest a biological basis for the "laziness" ascribed to them in earlier essays, which brings motivational characteristics into the realm of biologically fixed racial differences. In the end, Kant does not strictly commit himself to race differences with moral implications as profound as those of sex, but even insofar as he approaches these sorts of views, Kant's moral philosophy and philosophy of history give decisive reasons to favor treating Kant's characterizations of various races in just the way that Kant suggests in *Observations*, as a catalogue of traits to which individual members of those races can and should make exceptions of themselves. Even this way of reading Kant is hardly without its dangers. Members of racial groups who are classified as having particular defects can come either to be demoralized through Kant's theory or to overcompensate, trying to "prove themselves" in ways that go far beyond the actual demands of moral and cultural life. Even interpreted in the most generous way, Kant's race theory brings problems, and taken too far, it could license the worst racist abuses, as Kant's seeming support of slavery in the footnote above might forebode.

On the second issue – the relationship between Kant’s race theory and his *Critique of Judgment* – it would be nice to say, as we did with Kant’s informal observations about race, that one can insulate his Critical philosophy from his race theory. In fact, however, as Robert Bernasconi has rightly pointed out, “Kant’s understanding of race is at stake in the discussion of teleology in the *Critique of Judgment*” (Bernasconi 2002:147, see too Bernasconi 2001). But the relationship between race theory, teleology, and the central aspects of Kant’s Critical philosophy (and with it, his transcendental anthropology), is not as problematic as some commentators (e.g. Eze 1994) have suggested because it tends to be unidirectional. Race theory supports for Kant’s philosophy of biology primarily by showing a particularly interesting way in which it can be applied, but the general points that Kant makes use of in his race theory – such as the distinction between natural history and mere description of nature (2:434n, 8:153f) or the legitimacy of teleological principles in biology (8:157-84) – are, as Kant recognizes, compatible with different specific empirical accounts depending upon the empirical details to which they are applied. Where Kant’s race theories go astray is not at the level of these general methodological principles but in the specific and misguided application of them. Thus while problems with the accounts of judgment and teleology in the *Critique of Judgment* could undermine Kant’s race theory, abandoning his race theory altogether, if done for empirical and not methodological reasons, would not jeopardize his *Critique* at all.¹³⁴ Kant’s more general philosophy of biology does not *preclude* scientific racism, which may be an indictment of a sort, but it also does not *imply* Kant’s theory of race.

The final issue – the role of Kant in the development of modern racism – is in some respects the most complicated. Kant’s race theory, with its defense of monogenesis and emphasis on physical rather than moral or intellectual characteristics, is the sort of theory that one would expect a thoughtful, cosmopolitan humanitarian to develop in the 18th century. As a way of making sense of the confusing array of anthropological discoveries faced by Europeans coming into greater contact with the rest of the world, Kant’s race theory might have seemed to be extremely well suited to his moral ideals. All human beings are a single species and hence – one would think – equally worthy of respect and capable of the highest human ideals. But there are real, hereditary,

¹³⁴ Similarly, Kant’s theory of race could prove an important supplement to Kant’s moral philosophy. It might affect how one applied moral principle to other races (though see next section), or – more likely – it might affect the sorts of moral education we apply to different races. In either case, changes to Kant’s race theory could affect the *application* of his moral theory, but need not pose problems for the core of that moral theory.

biological differences between people that manifest in their physical appearance. So far, so good.

But Kant's race theory did not *in fact* limit itself to physical characteristics. We have already seen this in Kant's own life, where he combined a scientific account of race with intensely negative characterization of non-European races. But the same held true for those who appropriated the concept of race as a scientific concept in the 19th and 20th centuries. Given the role that "scientific racism" came to play in the entrenchment of racist ideology (especially in Europe and the Americas),¹³⁵ Kant's role in making this science possible implicates him in those racist ideologies, even if only indirectly. Kant's concept of race was not, of course, the most important influence on racism (even scientific racism) in the 20th century; Darwin's account was more important. And if Kant's personal views were different, his overall place in the history of racism might well be that of the well-intentioned but naïve humanitarian attempting to combat proto-racist tendencies through science, whose science ended up working to promote racism. But given Kant's personal views about race and the fact that his race theory *did* in fact end up playing a real role in the development of what has come to be modern-day racism, Kant must – unfortunately – be given a prominent place in the history of racism, as he has in the histories of human rights, aesthetics, theology, and other fields where his impact has been more positive.

c) Political issues: Slavery and colonialism

When Kant turns from *descriptions* of different races to *prescriptions* for how to deal with members of other races, the merits of his moral philosophy overwhelm the offense of his empirical anthropology. The two most important issues facing Europeans in their interactions with other races were the closely connected issues of colonialism and race-based slavery. And Kant's later political writings involve detailed and impassioned rejections of both. With respect to slavery, Kant's published writings are clear and direct.¹³⁶ He refers to

¹³⁵ See especially Gould 1981.

¹³⁶ Kant's unpublished writings are less clear about the issue of slavery. In notes written during the 1780s, Kant seems to accept the notion that "Americans and negroes cannot govern themselves. Thus [they] are good only as slaves" (15:878). Strictly speaking, these comments need not morally justify chattel slavery, but they seem at least in tension with Kant's later works. On the other side, Kant's notes for his *Perpetual Peace* include condemnations of slavery – especially the enslavements of non-whites by whites – that are

West Indian (or “Sugar Islands”) slavery as “the cruelest and most calculated slavery” (8:359) and insists in his *Metaphysics of Morals* not only that no one may sell himself into slavery (6:270) but also that *any* relationship between master and servant can be “at most only for an unspecified time, within which one party may give the other notice” and “children . . . are at all times free” (6:283). Whatever forms of indentured servitude might be allowed, the chattel slavery associated with the slave trade and European (especially British) colonialism are excluded. With respect to colonialism in general, Kant is even more eloquent:

If one compares with [the duty to universal hospitality] the *inhospitable* behavior of civilized, especially commercial, states in our part of the world, the injustice they show in *visiting* foreign lands and people (which with them is tantamount to conquering them) goes to horrifying lengths. When American, the negro countries, the Spice Islands, the Cape, and so forth were discovered, they were, to them, countries belonging to no one, since they counted the inhabitants as nothing. In the East Indies (Hindustan), they brought in foreign soldiers under the pretext of merely proposing to set up trading posts, but [brought] with them the oppression of the inhabitants, incitement of the various Indian states to widespread wars, famine, rebellions, treachery, and the whole litany of troubles that oppress the human race. (8:358-9)¹³⁷

Kant not only vehemently rejects the practice of “counting [non-white peoples] as nothing,” but he shows a surprisingly degree of insight into the deceptive justifications for standard practices of European colonialism, even emphasizing that his “right to hospitality – that is, the

even more impassioned and detailed than those in his published writings. Pauline Kleingeld summarizes some of these:

[Kant] sharply criticizes ‘the civilized countries bordering the seas’, whom he accuses of recognizing no normative constraints in their behaviour towards people on other continents and of regarding the ‘possessions and even the person of the stranger as a loot given to them by Nature’. Kant censures the slave trade (‘trade in Negroes’), not as an excessive form of an otherwise acceptable institution, but as in itself a ‘violation’ of the cosmopolitan right of blacks (23:174). Similarly, he criticizes the fact that the inhabitants of America were . . . ‘displaced or enslaved’ soon after Europeans reached the continent (23:173-4). (Kleingeld 2007: 15).

The combination of these comments lends credence to the view (noted below) that Kant changes his views on race in the 1790s.

¹³⁷ Kant’s quotation goes on to refer to the wisdom of the “great empire” of China, as well as Japan, in limiting access of Europeans to their lands (8:359).

authorization of the foreign newcomer – does not extend beyond the condition which makes it possible to *seek* commerce with the old inhabitants” (8:358), drawing attention to the practice of using trade to control other peoples. In his *Metaphysics of Morals*, Kant returns to the issue of colonialism, now connecting it with the same end of Nature that played such an important role in his race theory, the spreading of human beings across the globe. Kant claims that “all nations stand *originally* in a community of land” but emphasizes that this does *not* give anyone the right to possess any land they find. Instead, Kant emphasizes the right of first possession (6:263) and specifically applies this to the case of “newly discovered lands”:

If the settlement [of these lands] is made so far from where [the local] people reside that there is no encroachment on anyone’s use of his land, the right to settle is not open to doubt. But if these people are shepherds or hunters (like the Hottentots, the Tungusi, or most of the American Indian nations) who depend for their sustenance on great open regions, this settlement may not take place by force but only by contract, and indeed by a contract that does not take advantage of the ignorance of those inhabitants with respect to ceding their lands. This is true despite . . . it [being] to the world’s advantage . . . [A]ll these supposedly good intentions cannot wash away the stain of injustice in the means used for them. (6:353)

Again, Kant not only rejects the practice of seizing land in the “New” World, but specifically addresses two of the main ways in which this practice is justified, through appeals to the greater good (e.g., that American Indian land can be better used for farming than for hunting) and through spurious “contracts” whereby natives are robbed of their land by virtue of not fully understanding the costs and benefits at stake. Despite whatever personal views he may hold about the capacities of non-white peoples, Kant has no tolerance for failing to afford them the rights consistent with the humanity that we all share in common, and he eloquently defends those rights against encroaching colonial practices of Europeans.

There are at least two important ways of reading Kant’s moral and political stances towards other races, both of which at least partially redeem Kant as a thinker about race. First, one might rightly point out that Kant’s moral theory emphasizes the importance of respect for the humanity of others, where humanity primarily involves the mere capacity for choice. Just as Kant insists that a person’s (radical) evil does not

justify denying them respect, so he should equally insist that whatever differences there are between Europeans and non-Europeans, as long as non-Europeans are human beings with a capacity for choice, they must be respected. In that sense, Kant's moral arguments against slavery and colonialism are not only consistent with his overall moral theory but show the power of that moral theory even in the face of extreme personal prejudice. Second, one might find in Kant's political writings some hint that Kant's views on other races changed in response to criticisms by Forster, Herder, and others (see Kleingeld 2007). The comments in which Kant is disparaging of other races gradually cease after 1792, such that Kant's *Anthropology* (published in 1798), while in other respects following the general outline of his *Observations*, refrains from making any mention of his views on race, instead referring readers to a text by another author. This has led at least one scholar to argue that "Kant changed and improved his position[on race] during the 1790s" (Kleingeld 2007:3).

Neither of these ways of reading Kant's later work justifies the racially offensive comments found elsewhere. The fact that in his later years Kant did not use negative characterizations of other races to justify slavery or colonialism does not take away the real harm of depicting them as naturally inferior to whites, and Kant was well aware of the damage to one's personhood that comes from damaging the way that one is perceived in the eyes of others. Even changing his views eventually does not justify Kant's holding them for as long as he did. But just as Kant's other writings on race show the limits of Kant's moral philosophy, these later texts show some of its power. At the very least, the emphasis on universal human dignity in his moral philosophy helped prevent Kant from drawing the worst practical implications of his early views on other races. At best, his philosophy may have even helped him see the errors of those views.

V. Conclusion

Kant's transcendental and even empirical anthropology has often been criticized for being excessively universal, for painting all humans with the same brush and ignoring the differences between them, for being insufficiently attentive to the idiosyncratic, masculine, or European perspective from which Kant writes. Kant himself was deeply attuned to these concerns. Even before he started developing an a priori transcendental anthropology, Kant insisted on teaching a course in physical geography to "make good [his students] lack of experience" and

equip them to function in the world. An important part of this course was exposing his students to the variety of manners and types of people in the world, and from his early *Observations* to his eventually courses (and book) in anthropology, Kant expanded this part of his geography course to include detailed accounts of human difference so his students could act and judge well in a diverse human world.

In broad outline, this attention to human difference as an important part of being an active world citizen is an admirable aspect of Kant's overall anthropology, something that should make that anthropology more appealing today. But the details of Kant's accounts of diversity reveal a darker side to attending to human difference. Kant's resources for understanding human difference were limited. He never married and his primary knowledge of women was through formal dinner parties and English novels. He never traveled, so his primary knowledge of non-Europeans (and even most Europeans) was from books written by others. Kant made the most of books, voraciously reading travel logs, novels, scientific and medical treatises, and other accounts of human difference. But even with this background, Kant recognized that ultimately "the choice of the philosopher" plays a significant role in how such data is interpreted. And Kant's choices, with respect to both women and non-Europeans, were generally reprehensible. What could have been a significant *improvement* to Kant's anthropology now serves as a *warning* about the dangers of supplementing a transcendental anthropology with an empirical one, and especially of supplementing a universal anthropology with an account of diversity. Despite its errors and dangers, however, Kant's account of human diversity is an important part of his overall account of human beings, and his insistence that human difference as well as human similarity is an important part of what makes us who we are is an insistence that continues to resonate today (as we will see especially in chapter nine).

CHAPTER 5: PRAGMATIC ANTHROPOLOGY

The previous four chapters examined Kant's transcendental and empirical anthropologies, his normative and descriptive accounts of human life from-within and from-without, as these "anthropologies" are expressed in various written works and unpublished lectures composed by Kant over his lifetime. At the end of his life, however, when Kant came to write the only published work he entitled *Anthropology*, this book was neither a *Transcendental Anthropology* nor an *Empirical Anthropology*. Instead, Kant wrote *Anthropology from a Pragmatic Point of View*. And throughout his life, Kant taught "Anthropology" courses that cannot be described as either transcendental or empirical. Instead, what Kant sought to do, throughout his life and especially in *Anthropology from a Pragmatic Point of View*, was to develop a *new* approach to thinking about human beings, one that combined theoretical insight into human nature with humans' fundamental practical concerns, and one that avoided stale, metaphysical debates about such things as the relationship between mind and body, while providing a useful, philosophically sophisticated, systematic answer to the question "What is the human being?" In his "pragmatic" anthropology, Kant pulls together his transcendental and empirical anthropologies into a coherent whole that can help his readers actually "properly fulfill [our] station in creation" (20:41). Pragmatic anthropology thus marks a sort of culmination of Kant's anthropology. While not the arena within which Kant answers all the questions of philosophy, it is where he most fully combines philosophical insights with empirical-psychological observations of human beings in order to *improve* humans' cognition, feelings, and actions.

Kant does, of course, *discuss* "anthropology" in his previous published works, most famously in his *Groundwork* and *Metaphysics of Morals*. In the first, he explains that "ethics" will have an "empirical part" called "practical anthropology," which will help him "distinguish in what cases [moral laws] are applicable and . . . provide them with access to the will of the human being" (4:388-9) and "would deal . . . with the subjective conditions in human nature that help or hinder in fulfilling the laws of a metaphysics of morals" (6:217). Moreover, as we saw in chapter three, Kant's discussion of humans' radical evil requires some explanation for how one can work to undo and arm oneself against self-wrought evil tendencies, an explanation that one would expect from his moral anthropology. One might think, then, that Kant's *Anthropology from a Pragmatic Point of View* would provide the much-needed supplement for his pure moral philosophy.

In fact, *Anthropology from a Pragmatic Point of View* does provide this supplement, but the text also goes much further than a merely moral anthropology. From Kant's transcendental anthropology, one might think that Kant cares only about those aspects of human nature that are subject to a priori normative principles. And while his empirical anthropology shows his interest in *investigating* a much broader range of human action, one might be left with the impression that anything that does not ultimately serve moral, epistemic, or aesthetic ends is, for Kant, a flaw in human nature. Kant's *pragmatic* anthropology alleviates these concerns by showing Kant's interest in helping humans become better at *all* of those aspects of human life that make it worth living. As he explained in an early (1773) letter to a former student, his course in anthropology will "disclose the basis of . . . everything that pertains to the practical" (10:146, see too 16:804).

This general understanding of pragmatic as practical shows up early in Kant's published *Anthropology*. Kant distinguishes his "pragmatic" anthropology from a "physiological" one on the grounds that whereas the latter emphasizes "what nature makes of the human being," pragmatic anthropology attends to "what *he* as a free-acting being makes . . . or can and should make of himself" (7:119). More specifically, while the physiological anthropologist looks at the neurophysical bases of mental powers, "he must admit that in this play of his representations he is a mere observer and must let nature run its course, for he does not understand how to put [these neurophysical bases] to use for his purposes" (7:119). In contrast, the pragmatic anthropologist focuses on that "knowledge of the human being" that is *useful* for, say, improving memory; this anthropologist "uses perceptions concerning what has been found to hinder or stimulate memory in order to enlarge it or make it agile" (7:119).

Kant's particular example of memory is important for clarifying what exactly is meant when Kant refers to the subject matter of pragmatic anthropology as something that "concerns . . . the investigation of what he as a free-acting being makes of himself, or can and should make of himself." In particular, Kant does not imply here that pragmatic anthropology takes the free subject of transcendental anthropology as its primary *topic* of investigation. In the rest of Kant's *Anthropology*, not only through his stated methodology (7:120-22) but in his actual practice, the information about human beings that makes up the content of pragmatic anthropology is empirical. In an important sense, then, pragmatic anthropology has no distinctive content of its own. Elsewhere, Kant even denies a distinctive content to "practical psychology" – using the term psychology here to refer to what he

elsewhere calls anthropology – saying instead that all of its principles are mere “scholia” of empirical “knowledge of [human] nature” (20:199). Pragmatic anthropology is concerned with free-acting human beings in that it is addressed *to* human agents who can *make use* of empirical knowledge for accomplishing their (freely-chosen) goals.

When one asks where the ends served by pragmatic anthropology arise, one must look to transcendental anthropology in the broadest sense. It is only from-within the standpoints of thinking, feeling, and willing that one comes to discover – at least in outline – what the human being *should* make of himself. As freely-acting beings, human beings find themselves with two main orienting principles of volition: happiness and duty. And from the start, Kant recognized that these orienting principles require considerable empirical knowledge for their application. Happiness is “such an indeterminate concept that although every human being wishes to attain this, he can still never say determinately and consistently with himself what he really wishes and wills” (4:418) and even moral laws require “a judgment sharpened by experience . . . to distinguish in what cases they are applicable and . . . to provide them with access to the will of the human being” (4:389). Kant’s pragmatic anthropology thus fills in the empirical knowledge of human nature that is required in order to discern not only the means for pursuing one’s ends but even what ends Kant thinks free but finite human agents should devote attention towards. Of these, of course, the only end that is “good without limitation” (4:393) is the good will, so I begin my discussion of Kant’s pragmatic anthropology with more specifically *moral* anthropology.

1. Moral Anthropology

For Kant, “moral anthropology” provides a necessary empirical supplement to his “pure” moral philosophy, rooted in his transcendental anthropology of desire. Precisely what moral anthropology contributes is disputed, however, and even seems to shift between Kant’s earlier and later works in moral philosophy.¹³⁸ In particular, we can distinguish two different senses in which empirical anthropology might help supplement a pure moral philosophy. First, and most obviously, we might need what Kant, in *Groundwork*, refers to as a “judgment sharpened by experience . . . in order to distinguish in what cases [moral principles] are applicable” (4:389). Even for generating relatively specific principles from the

¹³⁸ By “earlier” and “later” here I refer to Kant’s *Groundwork* (1785) and *Metaphysics of Morals* (1797). In Kant’s earliest works in moral philosophy, such as his *Observations on the Feeling of the Beautiful and Sublime*, there is no distinction between (empirical) anthropology and (pure) moral philosophy.

categorical imperative itself, we might need to include various sorts of empirical information about human beings. For example, the obligation not to make false promises only applies to beings like ourselves, not, for instance, to the extra-terrestrials that Kant imagines in his anthropology, “rational beings on some other planet who could not think in any other way but aloud” (7:332). We might also use empirical facts about the human desires to specify and apply our general obligation to promote the happiness of others, and we can use facts about the limits of human capabilities to set proper boundaries and conditions for a wide variety of human interactions. As we will see at the end of the next section, Kant does see pragmatic anthropology as playing a very important role in this sort of application of moral principles to human life. But this role is *not*, at least not in his most developed accounts of morality, the role that he ascribes to “moral anthropology” strictly speaking.

Instead, “moral anthropology” is concerned with a second way of putting empirical anthropology to use for moral philosophy: “Moral anthropology . . . deal[s] with the subjective conditions in human nature that help or hinder in fulfilling the laws of a metaphysics of morals. It . . . deal[s] with the development, spreading, and strengthening of moral principles” (6:217). As Kant explained in lectures in both ethics and anthropology, empirical insight is needed to “make human beings ready to follow [moral laws]” (27: 244) and even to “give duties the power of inclinations” (25:1437, cf. 25:471-2, 734-5). The general idea here is that one can use empirical knowledge about how human volition actually works in order to help human beings – both oneself and others – act on the basis of the moral law.

This task of putting empirical knowledge to use for moral improvement is particularly important, for Kant, because of the problem of radical evil diagnosed in chapter three. There we saw that human beings have a self-wrought propensity to evil by virtue of which we not only perform evil acts but also cultivate in ourselves a tendency to do more and more evil. In response to the problem that human evil poses for his moral philosophy in that it seems to make a good will impossible for human beings, Kant proposes both a theological solution – God’s grace – and a set of practical solutions. In chapter three, we emphasized two aspects of Kant’s practical solution, his emphasis on human *historicity* and on the *ethical community*. But both of these aspects are supplemented by Kant’s emphasis on moral anthropology. Given our radical evil, the best that human beings can do with respect to morality is to “remain forever armed for battle,” “under the leadership of the good principle, against the attacks of the evil principle” that is in him (6:93). But this involves an unending effort to *strengthen* this good principle and

weaken the good principle in one's empirical character, an effort that can only be facilitated through empirical, anthropological knowledge of how human volition actually works.

For that reason, Kant develops a moral anthropology, an application of his empirical anthropology to the specific problem of how to cultivate virtue in human agents. His answers to this question stretch over several his works. Throughout his works in moral philosophy, Kant emphasizes the *practical-pedagogical* importance of presenting the moral law in all its purity (e.g. 4:390, 5: 156), and not only offers a “moral catechism” (6:480) in his *Metaphysics of Morals*, but – much more powerfully – sketches a model of how to use stories of moral heroes to bring a “ten-year old boy . . . from approval to admiration, from that to amazement, and finally to a lively wish that he himself could be such a [virtuous] man” (5:155-6). Kant's lectures on pedagogy include a long discussion of how to cultivate, from infancy to adulthood, the moderation and self-discipline that will provide a foundation for moral virtue. Kant uses, for example, his empirical claims about the nature of habit-formation and the role of inclinations in moral temptation to argue in these lectures that “the child should be prevented from getting accustomed to anything; he must not develop habits” (9:463). And Kant's lectures on ethics include attention to how “a person may be compelled to duty by others” (27:521) through the careful articulation of one's moral responsibilities. Unsurprisingly, then, Kant's *Anthropology from a Pragmatic Point of View* includes substantial attention to empirical features of human beings that are particularly salient for the cultivation of moral virtue and even ends with an impassioned reiteration of the moral vocation of human beings as an historical and social species (see 7:330-3). Four particular examples from this work are particularly important: politeness as an aid to virtue, passions and affects as hindrances to virtue, character as a necessary ground for virtue, and understanding diversity to cultivate morals properly.

Kant discusses politeness primarily in two contexts, in his *Metaphysics of Morals* where he insists on a “duty to oneself as well as to others” to engage in polite society (6:473), and in his *Anthropology* and related lectures, where politeness shows up as an example of a “permissible moral illusion” (7:151, cf. 25:502-505, 1455). In both contexts, Kant's emphasis is on the moral-anthropological importance of politeness, that is, on the way in which it either strengthens or weakens commitment to the moral law. Against philosophers like Hume and Smith, Kant emphasizes how politeness is significantly different in kind from truly moral action; it is a “mere external . . . which give[s] a beautiful illusion [merely] resembling virtue” (6:473). But against

Rousseau, Kant insists that this merely external show of virtue can, over time, “promote a virtuous disposition” (6:474). In his *Metaphysics of Morals*, Kant’s account of precisely *how* politeness does this is absent, but his *Anthropology* makes use of empirical facts about human beings – such as the “tendency to willingly allow himself to be deceived” that “nature wisely implanted” in human beings – to show how engaging in shows of virtue can, over time, help a person combat the evil principle within them and cultivate genuine virtue.¹³⁹

While politeness is an aid to moral character, passions and affects are the prime examples of moral hindrances in Kant’s *Anthropology*. As we saw in the last chapter, Kant treats affects and passions as “illness[es] of mind” (7:251), and both represent degrees of emotional agitation that rise to the level where one is incapable of reflecting on one’s goals. Both affects and passions preclude acting from a good will, since the reflection necessary to be motivated by the moral law is absent, but the way in which they preclude the good will is different. Affects are an extreme and passing form of weak or frail will, where one is simply overcome by feeling and engages in no reflection at all. Sudden rage or shock fall into this category. Fortunately, affects are passing, and one afflicted by affect can, during cool, calm hours, take steps to prevent further outbursts. Those afflicted by passions akin to depraved wills but focused, not on their own happiness overall, but solely on a single end. Vengeful hatred and the lust for power are among the passions. Kant uses his taxonomy of mental faculties – especially the distinction between feeling and desire – and his overall empirical account of human action to explain not only the difference between these mental illness but also why passions are not only “incurable because the sick person does not want to be cured” (7:266) but also, and in contrast to affects, “properly evil” (6:408).

Throughout his lectures and in his published *Anthropology*, Kant follows his discussion of affects and passions with an account of “character” that pulls together his empirical treatment of character with his moral philosophy. Character – the tendency to act in accordance with consistent principles – is the surest antidote to passions, affects, and even the ordinary human emotional fluctuations that can be so problematic to moral life. Thus Kant’s *Anthropology* emphasizes the importance of character for living a good life. Here it is important to recall that Kant’s empirical psychology emphasizes “character” as the proper expression of the higher faculty of desire, but also notes that most individuals have a “bad” or “flawed” character that fails to fully regulate

¹³⁹ For a detailed discussion of politeness in Kant, see Frierson 2005.

itself by consistent principles. Moreover, recall that radical evil in Kant's *Religion* is not diabolical – a rebelling against morality as such – nor even primarily depraved – a consistent preferring of happiness to morality – but rather an impurity or frailness that manifests itself in a tendency to give in to sufficiently strong temptations. For Kant, character – the “property of the will by which the subject binds himself to definite practical principles” (7:292) – is necessary in order to have a will that is stably and purely bound to the moral law as its principle. Thus Kant's *Anthropology* devotes considerable attention to cataloging measures that one can take to cultivate genuine character.¹⁴⁰

In the course of his discussions of individual character – the sort so prominent in his empirical theory of the higher faculty of desire – Kant also discusses other sorts of “character,” including the character of different nations and races, and various “influences on character,” including temperament. In other words, Kant's moral anthropology, especially when it comes to the matter of character, absorbs his interest in diversity. And in this context, his treatments of diversity take on a special sort of moral importance. In particular, Kant often emphasizes human differences as a way of highlighting the very different sorts of “struggles against the evil principle” that different people will have. In describing the sanguine as “not . . . evil . . . but . . . a sinner hard to convert [because] he regrets . . . much but quickly forgets this regret” (7:288) or the phlegmatic as one who “proceeds from principles” even when lacking “wisdom” (7:290), Kant highlights different emphases for moral cultivation in these two sorts of people. And Kant's claims that French “vivacity is not sufficiently kept in check by considered principles” (7:313) or that Germans “have a tendency to imitation . . .[,] a

¹⁴⁰ Of course, nothing in Kant's anthropology dictates empirical causes of a good will since a good will cannot be empirically caused. But Kant does lay out empirical causes of and practical advice for cultivating character, he highlights particular moral challenges that will be faced by those of varying temperaments and national origin, and he describes the resources for progress in the species as a whole. Because all of these considerations are empirical, they cannot ultimately determine whether one's free, noumenal choice is for good or evil. But a person who recognizes radical evil and earnestly seeks every resource to combat the “evil principle” within her knows that merely choosing rightly in a particular case is insufficient for moral progress. For radically evil human beings, virtue requires taking one's life as a whole to be a battlefield between one's evil tendencies to self-deceptive moral complacency and one's moral struggle to revolt against those tendencies. Knowing that one's friendliness or even principled actions are merely matters of temperament can help one focus on those areas of life that require greater moral attention (being more principled if one is naturally sanguine, or being less imitative if one is naturally phlegmatic). And given how lack of character leads to moral failings, knowledge of how to cultivate character, properly used, can arm one against the evils of frailty and impurity of will. Even if moral anthropology does not teach one how to effect a noumenal revolution against radical evil in one's will, it gives one the tools to effectively live out that revolution in a life of constant progress towards greater and greater conformity to the demands of the moral law. (For more discussion of these points, see Frierson 2003.)

mania for punctiliousness and . . . a need for methodical division . . . [that] reveals the limitation of [their] innate talent” (7:318-9) show that these two nations face different challenges in the cultivation of (moral) character (7:293). Kant’s pragmatic attention to empirical differences between human being often focuses on just those “subjective conditions” relevant for moral anthropology.

2. Pragmatic Anthropology

Pragmatic anthropology is not limited to merely *moral* anthropology but encompasses “everything that pertains to the practical” (10:146).¹⁴¹ And while Kant’s emphasis on the good will in his moral writings might lead some to think that, for Kant, there is little more to life than doing one’s duty, in fact Kant is adamant that human lives, to be really good lives, must include not only duty but also happiness and the increasing perfection of the whole range of predispositions with which nature has gifted us. Strikingly, in fact, while moral anthropology is an important part of Kant’s pragmatic anthropology, Kant consistently downplays the moral importance of his anthropological claims. In the passage where Kant describes, for example, the ways an inclination like idleness morally corrupts by making one think that “when one does nothing at all . . . he can do nothing evil” (7:152), his *emphasis* is on the fact that this idleness is the cause of “boredom,” a feeling of extreme dissatisfaction with oneself and one’s life (7:151). Similarly, politeness is important not merely for moral ends but for the sheer pleasure of polite company. Affects and passions are dangerous, and character beneficial, not merely for virtue but for happiness as well. Kant even discusses how an “*evil* character” can inspire “admiration” (7:293) and his main examples of the dangers of affects and passions involve the ways in which these prevent people from reflecting on “sum of all feelings of pleasure” (7:254) or the “sum of all inclinations” (7:265). And Kant’s account of diverse national characters is oriented towards discerning “what each can expect from the other and how each could use the other to his own advantage” (7:312). Anthropology is full of applications of Kant’s empirical anthropology towards helping human beings cultivate skills and capacities or become happier through better knowledge of human nature (both one’s own and those of others). For the purpose of this chapter, four examples of this

¹⁴¹ In his writings on moral philosophy, Kant even associates the “pragmatic” with guidelines aimed solely towards happiness (see 4:416-7), but we have already seen that “pragmatic anthropology” includes specifically moral anthropology as well as advice that is more “pragmatic” in this narrower, hedonic sense.

non-moral “pragmatic” anthropology suffice: memory, distraction, “the highest physical pleasure,” and “the highest ethico-physical good.”

Kant’s discussions of memory and distraction highlight the application of empirical anthropology to the perfection of capacities that are neither moral nor an obvious part of human happiness.¹⁴² With respect to memory, Kant objects to those who propose “impressing certain ideas on the memory by association with correlative ideas that in themselves . . . have no relationship at all with each other” on the grounds that “in order to grasp something in the memory more easily, we inconvenience it with still more correlative ideas,” which “is *absurd*” (7:183). Instead, Kant suggests, “*Judicious* memorizing is nothing other than memorizing, in thought, a *table* of the *divisions* of a system (for example, that of Linnaeus) where, if one should forget something, one can find it again through the enumeration of the parts that one has retained” (7:184). Kant adds, of course, various further mnemonic “tricks” such as “maxims in verse, since the rhythm has a regular syllabic stress that is a great advantage to the mechanism of memory” (7:184). Distraction, which Kant warns can lead to “forgetfulness” (7:185) or even “dementia” (7:207) if overused, can be effectively used as a sort of cleansing agent for the mind:

[O]ne can also *distract oneself*, that is, create a diversion for one’s involuntary reproductive power of imagination, as, for example, when the clergyman has delivered his memorized sermon and wants to prevent it from echoing in his head afterwards. This is a necessary and in part artificial precautionary procedure for our mental health. Continuous reflection on one and the same object leaves behind it a reverberation, so to speak (as when one and the very same piece of dance music that went on for a long time is still hummed by those returning from a festivity, or when children repeat incessantly one and the same of their kind of *bon mot*, especially when it has a rhythmic sound). Such a reverberation . . . molests the mind, and it can only be stopped by distraction and by applying attention to other objects; for example, reading newspapers. (7:207)

¹⁴² These examples also elegantly show the interplay of empirical-causal description with pragmatic purposes. In the case of distraction, for instance, Kant assumes that people are free in the sense that he directs this advice to someone who he takes to be capable of acting upon it. But the advice is based on a picture of human cognition that is determinist in the sense that it traces the causes of various changes in one’s cognitions, from the way in which continuous reflection on a single object causes “reverberation” to the ways that one can undo this reverberation by, for instance, reading newspapers.

While both good memory and the “mental health” referred to here may be helpful for happiness and/or virtue, Kant’s emphases in these passages is on putting empirical knowledge of human beings to use for the “pragmatic” purpose of perfecting cognitive powers *as such*, regardless of any moral or hedonic purposes to which these might be put.

From his long and detailed “pragmatic” discussion of cognitive faculty (which takes up more than half of his published *Anthropology*), Kant turns to the faculties of feeling and desire, and there devotes attention to “pragmatic” advice about how to make human beings happy. Those familiar with Kant’s moral philosophy may recall his periodic despair about the possibility of “imperative of prudence” even “presenting actions . . . as practically necessary,” even merely necessary for happiness (4:418, cf. 5:36).

[T]he concept of happiness is such an indeterminate concept that, although every human being wishes to attain this, he can still never say determinately and consistently with himself what he really wishes and wills. The cause of this is that all of the elements that belong to happiness are without exception empirical, that is, they must be borrowed from experience . . . Now it is impossible for the most insightful . . . finite being to frame for himself a determinate concept of what he really wills here. If he wills riches, how much anxiety, envy, and intrigue might he not bring upon himself . . . If he wills a great deal of cognition and insight, that might become only an eye all the more acute to show him . . . ills that are now concealed . . . If he wills a long life, who will guarantee him that it would not be a long misery? (4:418)

One might not expect a philosopher who so rails against the possibility of framing a determinate concept of happiness to offer empirically-rooted rules of prudence, but Kant does just this. His *Anthropology* lays out a general discussion of the nature of pleasure and pain, noting in particular that given the particular “animal life” that humans share, “pain must always precede every enjoyment” and “no enjoyment can immediately follow another” (7:231).¹⁴³ And Kant gives detailed analyses of inclinations, not primarily to show their moral dangers but to reveal how they are often self-defeating even from the standpoint of human happiness. Following Rousseau, Kant emphasizes that imagination and reason give humans capacities to generate desires that do not contribute

¹⁴³ For a detailed discussion of the shifts in Kant’s views that led him to these claims, see Shell 2003.

to lasting happiness. Even while discussing the cognitive faculty, Kant emphasizes implications of cognitive powers for happiness. For example, after distinguishing “attention” from “abstraction,” Kant notes, “Many human beings are unhappy because they cannot abstract. The suitor could make a good marriage if only he could overlook a wart on his beloved’s face, or a gap between her teeth . . . But this faculty of abstraction is a strength of mind that can only be acquired through practice” (7:131-2). Kant not only diagnoses a source of unhappiness but suggests a means for cultivating the cognitive power to have a happier life. The peak of his emphasis on happiness comes, however, at the of the first part of his *Anthropology*, where Kant offers specific accounts of “the highest physical good” and “the highest moral-physical good,” *both* of which, despite their use of the terms “good” and “moral,” are suggestions for how to best be *happy* given our human nature.

The “highest physical good” is that “greatest sensuous enjoyment . . . not accompanied by any admixture of loathing at all” found in “*resting after work*” (7:276). Kant identifies psychological features of human beings that interfere with this enjoyment, especially “laziness” (“the propensity to rest without having first worked” (7:276)) and he uses his account to explain phenomena as diverse as the appeal of “a game,” which is “the best distraction and relaxation after a long intellectual exertion,” the tendency of “a love story [to] always end with the wedding,” and the fact that “boredom will often affect us in such a manner that we feel driven to do something harmful to ourselves rather than nothing at all” (7:232-3). The “highest moral-physical good” is the way to unify “good living with virtue in *social intercourse*,” and it is found, for Kant, in “*a good meal in good company*” (7:277-8). Kant spends five full pages (7:277-82) detailing the importance of such dinner parties and how they should be conducted, including advice about the number of guests (ten), the order of conversation (first narration, then arguing, and finally jesting), and even the proper roles of “small . . . attacks on the [female] sex” (acceptable as long as they are “not shameful”) and dinner music (“the most tasteless absurdity that revelry ever contrived”).¹⁴⁴ Such dinner parties please by virtue of humans’ innate sociability, and they even channel our unsociability into conversation and “dispute . . . which stirs up the appetite” but which, because it avoids excessive seriousness, is consistent with “mutual respect and benevolence” (7:280-1). Unlike vain luxury and chasing after superiority over others, “the art of good living” is a skillfulness of choice in social enjoyment, which . . . mak[es] pleasure mutually beneficial, and is calculated to last” (7:250). The end

¹⁴⁴ For a detailed discussion of Kant’s account of dinner parties, see Cohen 2008.

result is social enjoyment consisting of a “stimulating play of thoughts” that “promotes sociability” and thereby “dresses virtue to advantage” (7:279-82). While seemingly “insignificant . . . compare[d] to pure moral laws,” the graces of “social good living” are not only enjoyable in themselves but also serve virtue by preventing it from becoming distorted into a cold “mortification of the flesh” (7:282), and the social interactions of a well-run dinner party help pave the way for polite and enjoyable moral exhortations amongst members of an ethical commonwealth. In the end, while the specifics of these discussions is important, far more important is the way that *Anthropology* enriches the austere conception of human life one might find in Kant’s transcendental anthropology with details that are neither moral, nor epistemic, nor aesthetic.

But why does Kant devote so much attention to how human beings can improve non-moral capacities and become happier? In part, one might appeal to moral incentives even here. As we noted at the beginning of the last section, one way in which one might use empirical anthropology is to specify the duties required by pure moral philosophy. And Kant’s *Metaphysics of Morals* explains that there are two “ends that are also duties . . . [:] *one’s own perfection* and *the happiness of others*” (6:386). Learning what makes people genuinely happy and what best facilitates the perfection of all “his natural powers . . . of spirit, mind, and body” (6:444) specifies, in empirically-informed and concrete ways, what one ought to *do* to work towards these two obligatory ends.

But Kant also emphasizes that human beings “unavoidably want” their own happiness (6:386, cf. 4:418, 5:25), and in *Religion*, Kant ascribes humans’ pursuit of happiness to “predispositions to animality . . . [and] humanity” that are, ultimately, parts of our more general “predisposition to good” (6:26). Even the perfection of our capacities – on which the majority of *Anthropology* focuses – is for Kant “a pragmatic purpose” (6:444) required because “these capacities “serve [one] and are given . . . for all sorts of possible purposes” (4:423). In contrast not only to monkish moralists who decry base animal desires for food and sex but even to moralists like Rousseau who decry the social inclinations that come from comparing ourselves with others, Kant endorses *as good* a wide range of human inclinations – for food, sex, social life, fine wine, good conversation, and so on. While not “good without limitation” (4:393), these other important even if “limited” goods in human life should be pursued as well as possible. So an important part of pragmatic anthropology is discerning, through careful empirical investigation, what sorts of activities best satisfy and delight human beings over the long term. The despair over rules of prudence that Kant seemed to express in his *Groundwork* is mitigated by a very serious effort throughout his

Anthropology to provide, practical advice that is as effective as any empirical investigation can be at providing for improvement and personal as well as social well-being.

3. *Empirical, transcendental, and pragmatic anthropology*

So far, this chapter has focused on various details of Kant's pragmatic anthropology, but it is finally time to ask how Kant saw his pragmatic anthropology in relation to his empirical and transcendental anthropologies. In other words, it is time to answer Kant's question – "What is the human being?" in an integrated way. As we have seen in this book, throughout his life Kant distinguished what I have called "transcendental" from "empirical" anthropology, particularly insisting that the latter not corrupt the former. For Kant, it is essential to work out normative demands a priori and from-within and to keep this investigation distinct from empirical claims about human beings. But Kant's pragmatic anthropology brings the two sorts of investigation together, which might seem to undermine both projects since it risks undermining empirical investigation with normative prejudices and corrupting transcendental theorizing with empirical description. Kant's explanations of what pragmatic anthropology consists in seem to exacerbate this tension. Kant refers to it as at once "knowledge of the world" that can even be broadened by "travel" (7:119-20) and "the investigation of what [the human being] as a free-acting being makes of himself, or can and should make of himself (7:119), seemingly pulling together empirical and transcendental anthropology in the most incoherent of ways. In fact, however, Kant's pragmatic anthropology offers a coherent and plausible model for how empirical and transcendental anthropologies should be integrated, and thus a model for what a complete "doctrine of the knowledge of the human being" (7:119) might look like.

Kant's claim that pragmatic anthropology attends to "what [the human being] as a free-acting being makes . . . or can and should make of himself" (7:119) comes in the context of his distinction between his own "pragmatic" anthropology from the "physiological" anthropology of his contemporary Ernst Platner. A merely physiological anthropology emphasizes "what nature makes of the human being" by examining neurophysiological bases of mental powers, so it "must admit that in this play of his representations he is a mere observer and must let nature run its course, for he does not understand how to put [these neurophysiological bases] to use for his purposes" (7:119). In contrast,

the pragmatic anthropologist focuses on “knowledge of the human being” with “the goal of applying this acquired knowledge . . . to . . . the human being because the human being is his own final end” (7:119). Similarly, in a handwritten personal note, Kant describes his approach to anthropology saying, “the historical kind of teaching is pragmatic, when it . . . is not merely for the school, but also for the world or ethics” (16:804). And in his lectures, he emphasizes that “anthropology is . . . indispensable and manages great uses” (25:1437). The general point of all these remarks is that Kant’s anthropology is meant to be pragmatic in the sense of practical rather than theoretical, something that not only gives knowledge about human beings, but that gives knowledge that one can put to *use*.

Importantly, Kant’s claim that pragmatic anthropology “concerns . . . the investigation of what *he* as a free-acting being makes of himself, or can and should make of himself” should not be taken to imply that pragmatic anthropology takes the free subject of transcendental anthropology as its primary *topic* of investigation. In the rest of Kant’s *Anthropology*, not only through his stated methodology (7:120-22) but in his actual practice, the information about human beings that makes up the content of pragmatic anthropology is empirical.¹⁴⁵ In an important sense, then, pragmatic anthropology has no distinctive content of its own. Elsewhere, Kant even seems to deny the possibility of distinctively pragmatic anthropology at all:

[Although one might offer] practical precepts, which concern the voluntary production of a certain state of mind in us (e.g. that of the stimulation or restraint of the imagination . . .)[, t]here is no practical [anthropology]¹⁴⁶ as a special part of the philosophy of human nature. For the principles of the possibility of its state by means of art must be borrowed from those of the possibility of our determinations from the constitution of our nature and, although the former consist of practical propositions, still they do not constitute a practical part of empirical psychology, because they do not have any special principles, but merely belong among its scholia. In general,

¹⁴⁵ For a detailed defense of this claim, see Frierson 2003. It is worth noting, of course, that even Kant’s empirical accounts of human beings treat them as free beings in the sense that humans have a faculty of desire capable of acting for the sake of ends and a *higher* faculty of desire capable of acting from principles. Consistent with his empirical anthropology, Kant’s pragmatic anthropology treats human beings as “free” in that sense (cf. Cohen 2009). But what makes the anthropology *pragmatic* is not the way humans are treated teleologically-biologically rather than mechanically, but the fact that Kant emphasizes those aspects of human nature that free human beings can *use*.

¹⁴⁶ Although he says “psychology” here, his use of the term psychology in this context corresponds to what I have been calling “empirical anthropology.”

practical propositions . . . always belong to the knowledge of nature and to the theoretical part of philosophy. (20:199)

The denial in passage of a distinctive, practical anthropology might seem to preclude any *Anthropology from a Pragmatic Point of View* at all. But in fact, the passage merely clarifies the logical status of such anthropology. Technically, *Anthropology from a Pragmatic Point of View* is a subset of Kant's empirical anthropology, a series of "scholia" that apply that empirical anthropology to specific human ends. But precisely because pragmatic anthropology is oriented towards ends, it refrains from discussing *merely* theoretical aspects of human nature, claims about human beings that *cannot* be put to any use.

In what sense, then, is pragmatic anthropology concerned with a free-acting human being? Put simply, pragmatic anthropology is knowledge of human beings *addressed to* human agents who can *make use* of such knowledge for accomplishing their (freely-chosen) goals. Kant uses the case of memory to explain that what makes an anthropologist pragmatic is that he "uses perceptions concerning what has been found to hinder or stimulate memory in order to enlarge it or make it agile" rather than dwelling over purely theoretical claims about memory that are "a pure waste of time" (7:119). And so when he turns to discuss memory in detail in *Anthropology*, Kant uses his empirical anthropology to discern the best mnemonic strategies, rather than merely explaining the "nature" of memory as such. This strategy is even clearer in the discussion of distraction cited in the previous section. Recall that Kant recommends distracting oneself to avoid reverberation in imagination (7:207). Throughout his discussion, Kant assumes that people are free in the sense that he directs this advice to someone capable of acting upon it. But the advice is based on a picture of human cognition that is strongly determinist in that it traces causes of changes in cognitions, from the way continuous reflection on a single object causes "reverberation" to the ways that one can undo this reverberation by, for instance, reading newspapers. Similarly, Kant's accounts of how politeness cultivates good character, the nature of affects and passions, and even the general role of character in action are all *empirical-anthropological* claims, but ones that can be put to use. Kant *addresses* free agents, teaching empirical facts about human nature in order to show what free *human* beings can make of themselves, and how.

But pragmatic anthropology is not *merely* empirical anthropology that can be put to use. Kant also, implicitly if not explicitly, recommends *how* one should put one's empirical knowledge to use. Pragmatic anthropology teaches both what human beings *can* make of themselves

and what they *should* make of themselves. And these *norms* cannot be justified merely empirically but depend upon the from-within perspective that is the focus of Kant's *transcendental* anthropology. The molestation of mind that calls for distraction appears only from-within, in one's response to one's cognitive state. And Kant's discussion of cognition even ends with a series of "unalterable commands" that "lead to wisdom," including the need "to think *for oneself*" and "Always to think *consistently*" (7:228) and with an exhortation, fundamental to the autonomy at the core of Kant's transcendental anthropology, to "exit from his self-incurred immaturity," to "venture to advance, though still shakily, with his own feet on the ground of experience" (7:229). By providing free human thinkers with empirical knowledge about how cognition works, Kant can cultivate the autonomy of thought that is a normative requirement of thinking, evident from-within. Similarly, as we saw in section one above, Kant's pragmatic anthropology puts empirical anthropology to use to cultivate *good character* in human beings. By understanding the aspects of our natures that tempt and lead us astray and by understanding how to cultivate character, we can better engage in the struggle towards the good will that Kant's transcendental anthropology of desire shows, from-within, to be the only thing good without limitation. Pragmatic anthropology thus unifies transcendental with empirical anthropology; transcendental analysis provides the a priori normative principles for our human powers, and empirical anthropology shows how to cultivate powers that conform to those norms.

One aspect of Kant's pragmatic anthropology, however, goes beyond the mere application of empirically-given means to transcendently-given ends. Kant's discussions of happiness, which play a particularly important role in his "pragmatic" anthropology, make use of empirical anthropology to specify the best ends for human beings. Happiness is a universal end, one that we can observe all humans seeking and that we can discover, from-within, to be a naturally necessary end for ourselves. But happiness is "such an indeterminate concept" (4:418) that it requires empirical content to be action-guiding *at all*. Whereas there can be a priori cognitive and moral principles, the best that humans can do regarding the pursuit of happiness is to carefully study, through introspection as well as the investigation of others, what actually gives us the most pleasure over the long term.¹⁴⁷ In that sense, the pursuit of happiness is the special domain of pragmatic anthropology, the domain within which pragmatic anthropology specifies

¹⁴⁷ This should not be taken to deny that there can be an a priori investigation of pleasure, as chapter one makes clear. But this investigation only focuses on aesthetic – that is, *non-action-guiding* – pleasure.

not only means but also – because “happiness” is so indeterminate – the true nature of the end itself. Thus it is no surprise that Kant often uses the term “pragmatic” to refer specifically to what is to one’s own advantage, what is conducive to one’s own happiness (see, e.g., 4:416).

There is one further important aspect of Kant’s “pragmatic” anthropology that has not yet received sufficient attention. So far, this chapter focused on the importance of pragmatic anthropology for fostering *one’s own* capacities, happiness, and virtue. But for Kant, one of the primary reasons for developing a pragmatic anthropology is to learn how *others* respond to various empirical conditions in order to appropriately navigate within a world defined largely by other people. Thus Kant claims “all pragmatic instruction is instruction in prudence [*Klugheits Lehren*]” (25:471, see too 25:1436), in *Groundwork*, identifies “prudence” and “worldly wisdom” with “the skill of someone in influencing others so as to use them for his own purposes” (4: 416n.), and in *Anthropology* itself, defines the “pragmatic predisposition” as the ability “to use other human beings skillfully for his purposes” (7:322). Thus in laying out characteristics of different nationalities, Kant claims, “In an anthropology from a pragmatic point of view, . . . the only thing that matters to us is to present the character of [each] . . . in some examples, and, as far as possible, systematically; which makes it possible to judge what each can expect from the other and *how each could use the other* to his own advantage. (7:312, emphasis added). By understanding human nature, one not only knows how to influence *oneself* in order to improve memory, attain happiness, or cultivate character; one also knows or can quickly assess the strengths and weaknesses of others in order to influence their development and behavior.

This way of approaching others can sound sinister when Kant speaks of “using” others “to one’s own advantage,” but Kant’s point is more benign, and, properly understood, even an affirmation of human dignity. The fact is that human beings have various psychological characteristics that affect how we respond to each other. Regardless of how free one might be from-within, there are empirically-knowable tendencies in human nature that enable prediction of how different people respond to situations of various kinds. One *might* put this knowledge to use to manipulate others as *mere* means to one’s ends. But one might also put this knowledge to use in order to best achieve one’s ends *without* treating others as mere means. If I know that the sanguine “attributes a great importance to each thing for the moment, and the next moment may not give it another thought” (7:287-8) while the melancholy “finds cause for concern everywhere” (7:288), I know to treat

social commitments with these two types of people differently. If I have agreed to go to a movie with someone but feel like going to a concert instead, I will suggest this change to my sanguine friends (who are likely to be thrilled by my spontaneity, and in any case will tell me if they still want to go to the movie) but not to my melancholic ones (who are likely to silently take offense while reluctantly granting assent to the change of plans). Knowing how others are affected by empirical conditions can make me *more* respectful of their humanity, rather than less. Thus Kant says to his students,

We must trouble ourselves to shape the ways of thinking and powers of those with whom we have to act, so that we are neither too hard nor too offensive. So we are taught anthropology, which shows us how we can use people for our ends. (25:1436)

One uses anthropological knowledge, *rather than* manipulation or force, in order to accomplish one's ends in relation to others while still avoiding "offense" against their humanity.

Even this level of interest in others, of course, might still focus merely on *not* offending others while pursuing "one's own advantage." But Kant makes clear throughout his lectures on anthropology, and at the end of his published *Anthropology*, that his ambitions with respect to others are even higher. In one lecture, Kant laments that "the reason that morals and preaching that are full of admonitions . . . have little effect is the lack of knowledge of man" (25:471-72, cf. 27:358). In another, he explains:

Anthropology is . . . indispensable and manages great uses.

1. In pedagogy.

2. With respect to the influence that we can have on others. Especially commanders, who with a proper knowledge of man can get total opposites to work [together], which otherwise can be set right only through violence.

3. With respect to the influences on morals and religion, that through this knowledge one can give these duties the power of inclinations. (25: 1437)

And Kant's published *Anthropology* ends with an inspiring allusion back to the discussion of radical evil and the ethical commonwealth that Kant discussed in his *Religion*:

[Anthropology] presents the human species not as evil, but as a species of rational beings that strives among obstacles to rise out of evil in constant progress toward the good. In this its volition is generally good, but achievement is difficult because one cannot expect to reach the goal by the free agreement of *individuals*, but only by a progressive organization of citizens of the earth into and toward the species as a system that is cosmopolitically united. (7:333)

These three passages suggest a crucial *moral* function of pragmatic anthropology. Learning how other people are affected by empirical influences is crucial for “morals and preaching,” “education . . . [and] influences on morals,” and ultimately “a progressive organization of citizens are earth.” In these ways, we can see pragmatic anthropology as the science needed in order to actively promote an ethical commonwealth on earth. Thus it is unsurprising that at the outset of Kant’s *Anthropology*, he distinguishes several forms of “egoism” wherein one treats oneself as the ultimate authority in matters of truth, taste, or morals as hindrances to human excellence in these areas, and exhorts his readers to “the way of thinking in which one is not concerned with oneself as the whole world, but rather regards and conducts oneself as a mere citizen of the world” (7:130).

4. Conclusion

In the end, Kant’s *Anthropology from a Pragmatic Point of View* does, in a sense, provide his answer to the question “What is the human being?” For Kant, the question is a pressing practical one, a question about humans’ place in the universe, about who we are but also, crucially, about what we can and should make of ourselves. Many aspects of Kant’s answer to this question come in his transcendental anthropology. There he lays out the norms that should govern our cognition, feelings, and volitions; he shows the conditions of possibility of being bound by these norms; and he develops his metaphysical account of humans’ free and finite natures. Many aspects of Kant’s answer show up in his empirical anthropology, where he provides detailed descriptions of how human beings actually think, feel, and choose. But Kant’s pragmatic anthropology provides Kant’s way of integrating these other two sets of answers into a science that would help human beings better satisfy norms of cognition, feeling, and volition and live happier, more virtuous, and more beneficially-social lives.

This model of pragmatic anthropology helps complete Kant's accounts of human evil and human history, showing what we can do here and now, given the natures that we actually have, to make our world a better place. Moreover, as Kant's anthropology makes clear, human beings get into moral trouble both by prioritizing non-moral incentives over moral ones, and simply by being inconsistent, foolish, excessively passionate, and inept. The second problem can rise to the level of a genuinely moral problem, especially given our evil tendency to cultivate ineptness to excuse our failings. But it is also just a problem for living good lives. And Kant's pragmatic anthropology provides numerous specific suggestions for overcoming the general problems that make us both vicious and miserable. This approach also puts Kant's accounts of diversity in a new light, showing how they are meant to provide tools to *improve* humanity in all its forms, rather than excuses for dismissing or exploiting others.

This model of pragmatic anthropology also provides a useful model for how philosophical investigation of human nature might interact with human sciences today. While I discuss this in more detail in succeeding chapters, Kant's general model is to use transcendental philosophy to articulate and provide conditions of possibility for the norms governing human life and to use the findings of empirical sciences both to show the best ways of cultivating human beings who can achieve those norms and to discern and specify the one human end that is genuinely empirical: our pursuit of happiness. In the end, the answer to the question "What is the human being?" will be provided by philosophical accounts of the from-within, norm-governed perspectives of free and finite beings like us engaged in thinking, feeling, and choosing, along with empirical-scientific accounts of the characteristics and causal laws governing *homo sapiens*, combined into pragmatic knowledge that helps us become better-functioning, happier, more virtuous citizens of the world.

Chapter 6: The Immediate Reception of Kant's Anthropology

Kant had high hopes for his philosophy of the human being. He claimed that his Critiques, especially the Critique of Pure Reason, “completely specified [reason’s] questions according to principles, and, after discovering the point where reason has misunderstood itself, . . . resolved them to reason’s complete satisfaction” (Axii). He wrote of his course in anthropology that he would make this “very pleasant empirical study” into “a proper academic discipline” that could, “distinct from all other learning, . . . be called knowledge of the world” (10:146). Kant’s account of God, immortality, and human evil would, he hoped, put religion “within the boundaries of reason alone.” But while Kant did inaugurate a Copernican turn in philosophy, this turn did not take the form he expected. At almost every step of the way, its reception was disrupted in unanticipated ways. While Kant had hoped and expected that his transcendental philosophy would be sympathetically reviewed by prominent rationalist philosophers such as Moses Mendelssohn, the first reviews of the Critique of Pure Reason were dominated by empiricists and highly negative. Before discussion of Kant’s philosophy could reach its natural conclusion, it was overtaken by a controversy about Lessing that displaced the *Critique’s* detailed nuances in favor its “spirit.” In this context, a young upstart philosopher named Karl Leonhard Reinhold (1757-1823) emerged as the spokesman and heir of Kant’s philosophy – this even while Kant was still publishing major works of his own. Through Reinhold’s influence, the reception of Kant’s philosophy was driven along two radically different lines. German Romantics reacted against Reinhold’s attempt to transform Kant’s Critical philosophy into a foundationalist metaphysics by emphasizing inscrutability and particularity in a way that privileged artistic expression over philosophical system-building. German Idealists – Fichte, Schelling, and Hegel – took almost the opposite approach, building massive philosophical systems to fill Kant’s gaps and correct his errors. Meanwhile Kant’s empirical and pragmatic anthropology was largely ignored. By the time Kant published his *Anthropology from a Pragmatic Point of View* in 1798, the work barely attracted attention from a philosophical community that was preoccupied with the latest works by Reinhold, Fichte, and the early German Romantics.¹⁴⁸ By the time Kant died six years later, Schelling and Hegel were emerging as the most

¹⁴⁸ In 1798, the year Kant published his *Anthropology from a Pragmatic Point of View*, Reinhold published his *Fundamental Concepts and Principles of Ethics*, Fichte published *An Attempt at a New Presentation of the Wissenschaftslehre*, the Early Romantic Journal *The Athenaeum* was founded, and Schelling published *On the World Soul*.

important philosophers in Germany, and while both were heavily influenced by Reinhold and Fichte, they largely left Kant's own philosophy behind.

I. Early Reception and the Emergence of German Idealism

After Kant's *Critique of Pure Reason* was published in 1781, its first review (by Johann Feder and Christian Garve, in 1782) suggested that Kant was merely a new Berkeley who "transforms the world and ourselves into representations" (Sassen 2000:53).¹⁴⁹ Succeeding years brought a vibrant debate about Kant's philosophy, mostly characterized by empiricist criticisms of his philosophy and more-or-less Kantian responses to those criticisms. There were five key criticisms articulated during these early years. First, empiricists argued that Kant's Copernican turn implies an unacceptable Subjectivism, reducing all of our knowledge to mere knowledge of our own representations and precluding any real connection to an objective world. As one reviewer put it, Kantian sensibility "does not present the objective world at all" but is "like a glass whose exterior has an entirely foreign painting glued to it" (Pistorius 1786:115, in Sassen 2000: 101). Second, they argued that Kant has a Self-Reference Problem. If all that one knows is how things appear, and especially if self-knowledge is only knowledge of oneself as appearance, then "according to the author's system, appearance [cannot] be possible at all [since] that through which all appearing becomes possible (which must, accordingly, be presupposed prior to all appearances . . .) is supposed to be appearance" (Pistorius 1786:93, in Sassen 2000: 94). Put another way, the conditions of possibility of experience, and especially the transcendently unified "I" that is the subject of experience, cannot themselves be mere appearances. So Kant must be able to know something that goes beyond mere appearances, since he makes claims about the "I" that experiences those appearances. Third, empiricists raised the problem of the Neglected Alternative. Even if Kant can show that human beings necessarily impose sensory and cognitive structures on our experience of the world, nothing prevents the world itself from conforming to these categories. In fact, it seems reasonable to think that human beings have the cognitive capacities that we do precisely because these conform to the way that the world actually is. Fourth, empiricists raised a Problem of Affection. As F. H. Jacobi put it, "without the presupposition [of things in themselves], I could not find

¹⁴⁹ George Berkeley (1685-1753) was an Irish philosopher whose "idealism" advocated the view that "to be is to be perceived," that is, that only perceptions (and perceivers) are real.

my way into the system, whereas with it I could not stay there” (Sassen 2000: 173). Kant’s account of sensibility as a passive faculty seems to depend upon things-in-themselves affecting sensibility, but Kant’s limitations on knowledge make it “if not altogether impossible, at least quite difficult, to think things in themselves as the possible foundation or substratum of appearances” (Pistorius 1786:108, in Sassen 2000: 100). For Kant, the category of causation seems applicable only to appearances and thus not at all to the relationship between things-in-themselves and the sensibility that these things supposedly affect. Finally, more linguistically- and historically-inclined empiricists such as Hamann and Herder suggested that rather than Kant’s Copernican turn, something more like a Linguistic Turn was needed. Hamann critiques the “purism of pure reason,” Kant’s attempt to free reason from “tradition and custom” and from “language, the only, first, and last organon and criterion of reason” (Hamann 1800, in Schmidt 1996:155). He suggests that “space and time have made themselves universal and necessary” through the importance of “sounds and letters,” the “true aesthetic elements of human knowledge” that have their roots in “the oldest language . . . music, and . . . the oldest writing . . . painting and drawing” (Hamann 1800, in Schmidt 1996:156-7). Rather than space and time being fundamental a priori structures of human cognition, they reflect only the most general elements of human language. Hamann mocks the “Homer of pure reason” (Kant) as having “dreamed that the universal character of a philosophical language . . . is already found” (Hamann 1800, in Schmidt 1996:158).

These criticisms highlight several features of the early reception of Kant’s transcendental anthropology. First, reviews focused on the epistemic and metaphysical implications of Kant’s view, remaining faithful to the sense in which Kant’s first Critique emphasized the conditions of human *knowing*. Second and most importantly, these early reviews refused to endorse Kant’s Copernican turn. Rather than seeing the world of experience as largely a product of the particular structures of human cognitive powers, early reviews of Kant’s work insisted that cognition reflects rather than constructs the world it experiences. Because of this refusal to accept Kant’s Copernican turn, early reviews largely failed to grasp the force of Kant’s transcendental anthropology. The Self-Reference Problem, the Problem of Affection, and arguably even the Linguistic Turn all arise from trying to make Kant’s transcendental anthropology explanatory in the way that an empirical anthropology would be. Having portrayed human cognition as ultimately responsive to experience, the empiricist reviewers read Kant’s transcendental

anthropology as confused empirical psychology. Finally, early reviews reflect readers' general confusion about most details of Kant's arguments; they typically focus on overarching themes of Kant's work, and insofar as they enter into details, do so primarily in the context of the relatively straightforward accounts of sensibility rather than the very complicated transcendental deduction of the categories of the understanding.

After the first reviews of his Critique, Kant tried to correct misunderstandings and defend his transcendental anthropology, most notably through his publication of a simplified version of the main results of the Critique – his *Prolegomena to Any Future Metaphysics* (1783) – and his revision of the Critique itself (which aimed especially to distinguish his view from Berkeley's). But as Kant continued to develop his philosophy into ethics, aesthetics, and other areas, the work of defending and clarifying his first *Critique* largely fell to others.

Moreover, in 1785, after largely unfavorable initial reviews, an event transformed German intellectual life and contributed to a more favorable reception of Kant's philosophy. The event hardly seems remarkable: within a month, two biographies of Gotthold Ephraim Lessing, who had died a few years earlier, were published. Lessing had been the leading light of the German enlightenment for decades. (Kant had written, just a few years before publishing the Critique, "I am uncomfortable that you praise me as comparable to Lessing. For in fact I have not yet accomplished anything to deserve such comparison" (10:198).) The biographies were written by Moses Mendelssohn, another leading Enlightenment philosopher, and F.H. Jacobi, a long-standing critic of thoroughgoing Enlightenment rationalism. Both had known Lessing, and Jacobi had visited him during his final days. During those conversations, Jacobi reports that Lessing confessed, "There is no other philosophy than that of Spinoza" and that he adhered to the philosophy of "One and all."¹⁵⁰ In late 18th century Germany, "Spinoza" was a trope for materialism, fatalism, and atheism; Spinoza's deterministic pantheism – the view that everything is God and everything about God is necessary – was seen undermining both religion (since a philosophy in which God is identical to the world is no different from atheism) and morals (since if everything is necessary, then there is no room for moral responsibility). If Lessing, the best proponent of Enlightenment rationalism, had found that all philosophy ends in Spinozism, one seemed forced to choose between being rational and being moral-religious. Given this choice, Jacobi argued, one ought to subordinate

¹⁵⁰ See Lessing 2005: 241-56, especially p. 244.

reason to a “mortal leap” of faith. Mendelssohn, partly to save Lessing’s reputation but primarily to vindicate Enlightenment rationalism, sought to show that reason need not require abandoning morals and religion. The “Pantheism Controversy” that started with a dispute about Lessing soon became a wide-ranging debate about the prospects for Enlightenment rationalism in Germany.

The result of this apparently biographical dispute was that in 1785, German intellectual life became preoccupied with the issue of how (and whether) to reconcile faith and reason. And into that context came a new young philosopher and defender of Kant: Karl Leonhard Reinhold (1757-1823). In a series of “Letters on the Kantian Philosophy” published in 1786-7, Reinhold argued that Kant’s philosophy successfully solved this problem by negotiating a middle ground between “naturalists” who claim too much for reason and “supernaturalists” who allow too little. Reinhold, anticipating points that Kant made explicit in later works, suggested that even while justifying the possibility of a priori knowledge of objects of possible experience, Kant’s Critique of Pure Reason precludes rational knowledge of the all-important topics of religion: God and immortality. But in contrast to those like Jacobi who insist upon a “mortal leap,” Kant “found the true ground of conviction in a faith commanded by reason” (Reinhold 2006: 21). What’s more, by placing the ground for religious conviction in a “moral ground of cognition” (Reinhold 2006: 44), Kant protects against both amoral Spinozism and the moral-religious fanaticism that can come from non-rational religious faith. Kant’s distinction between theoretical and practical reason – or between the transcendental anthropologies of cognition and desire – provided a form of non-Spinozist rationalism to meet the needs of the age.

Reinhold’s Letters had three main effects on the reception of Kant’s philosophy. First, they solidified Kant’s status as the main philosopher to be reckoned with over the next thirty years. Kant’s work had been widely if critically reviewed, but the difficulty of its prose and the complexity of its arguments had prevented the widespread interest in Kant that came with Reinhold’s accessible and popular Letters. Second, because he situated Kant in the context of the Pantheism Controversy, Reinhold explicitly bracketed “the depths of speculation from which Kant has unearthed so many previously undiscovered treasures of the human spirit” in favor of “extracting the principle results of the Critique of Reason” (Reinhold 2006: 16). Thus Reinhold suggested that “his readers postpone their judgment on its internal grounds” (Reinhold 2006: 50n), and it is not until his last letter that Reinhold even begins to lay out some of the complicated transcendental anthropology of cognition that underlies Kant’s insistence upon rational faith (see Reinhold

2006:104ff.). While Kant's overall conclusions were seen as central to philosophy for the next several decades, the particular arguments used to establish those conclusions were largely left behind. Finally, the fact that Kant became popular through Reinhold quickly established Reinhold himself as a force to be reckoned with in German philosophy. Reinhold was given a Professorship at Jena within a year of publishing the Letters, and he quickly became seen – much to Kant's chagrin – as the defender of Kantian philosophy. Unfortunately for Kant, the third of these effects was to be the most lasting. Even as the tide turned from the practical and religious concerns of the 1780's back towards more systematic issues, philosophers increasingly turned to Reinhold rather than to Kant. By 1794, Fichte could write that “in the opinion of a majority of the admirers of the Kantian philosophy, Reinhold is the author who has either already succeeded in establishing philosophy as a science or else has best [i.e., better than Kant!] prepared the way for such success” (Review of *Aenesimius* 4, in Fichte 1993: 60).

While his early defense of Kant focused on the results of Kant's philosophy for the Pantheism Controversy, Reinhold quickly turned to fleshing out the “internal grounds” for Kant's transcendental anthropology in way that – he claimed – would be faithful to the “spirit” of Kant without becoming “slave to the letter” of Kant (Foundation 97, in diGiovanni 1985: 79). But Reinhold came to this project with commitments that were at odds with Kant's philosophy of the human being, which led to several revisions of Kant's transcendental anthropology: a shift towards foundationalism, a return to quasi-Leibnizian reductionism about mental states, a new kind of systematicity, and a non-anthropological approach to transcendental philosophy. First, Reinhold was a foundationalist in a way that Kant never was. Reinhold spends the decade after publishing the remarks in search of the “one absolutely fundamental explanation” that could irrefutably ground results he endorsed from Kant's Critique. Shortly after publishing his Letters on the Kantian Philosophy, Reinhold set to work on Concerning the Foundation of Elementary-Philosophy (1789), followed shortly thereafter by On the Foundation of Philosophical Knowledge, together with some comments concerning the Theory of the Faculty of Representation (1791) and Concerning the Foundations of Philosophical Knowledge, of Metaphysics, Morality, Moral Religion, and the Doctrine of Taste (1794). Whereas Kant sought to lay out transcendental conditions of possibility of various commonsense commitments (e.g., to empirical knowledge and moral responsibility), Reinhold aimed to lay indubitable foundations for philosophy in a basic starting point. As the subtitle of his 1791 text suggests, Reinhold located this foundation in the “faculty of

representation,” the definition of which he “does not pull out of thin air” but “ground[s] upon a principle which is determined through itself” (R81, in diGiovanni 1985: 71-2). This “principle” is “the principle of consciousness,” which is itself nothing more than “the exposition of the immediate expression of the self-explanatory actual fact of consciousness” (R103-4, diG 81). For Reinhold, this fact is “the only thing about whose actuality all philosophers agree . . . No idealist, no solipsist, no dogmatic skeptic, can deny its being” (R 190, cf. diGiovanni 1985:107-8).

Reinhold’s foundationalism contributed to a second key difference from Kant, a subtle rejection of Kant’s multi-faculty approach to the human mind in favor of reductionism about human mental states.¹⁵¹ In chapter three, we saw how Kant’s empirical psychology argued against attempts of Leibniz and Wolff to reduce all mental powers to a single power of representation. Similarly, in his transcendental anthropology, Kant carefully distinguished between the a priori structure governing different ways of approaching the world: as an object of knowledge, a sphere for human action, or a purposive locus of beauty and order. In his attempt to develop a systematic philosophy built on a single, solid foundation, however, Reinhold returned to the traditional Leibnizian strategy of reducing different mental powers to variations of a single power:

[T]he science of the a priori form of representing through sensibility, understanding, and reason; on this form depends the form of knowledge, as well as that of desire. In a word, it would be the science of the entire faculty of representation as such (R 71, diGiovanni 1985: 67)

[T]he two particular parts of the Philosophy of the Elements, the theoretical and the practical, of the sciences of the faculties of cognition and desire respectively, are subordinate to one common fundamental science” (Foundation 127, diGiovanni 1985: 91)

Reinhold’s foundationalist objection to Kant’s Critique of Pure Reason – that “the science of the faculty of cognition would have to be preceded by another that establishes its foundation” (R71, diGiovanni 1985: 67) – applies just as well to Kant’s other Critiques. And Reinhold’s supplementation of Kant’s specific Critiques with a foundation in

¹⁵¹ Admittedly, Reinhold did point out that different powers “express particular kinds of consciousness” that are irreducible to each other and that have “specific properties,” but his focus ends up being on the “principle of consciousness in general” that “expresses what is present in every consciousness” (Foundation 106, diGiovanni 1985:82, cf. diGiovanni 1985: 69, 70)

consciousness provides not only an indubitable ground for knowledge about our more specific mental faculties, but a ground that unifies these diverse faculties into merely different “specifications” of a single underlying faculty.¹⁵²

Third, Reinhold’s Leibnizian reductionism about human mental states combined with his strong foundationalism gave Reinhold’s philosophical anthropology a much more systematic character than Kant’s own. Kant’s anthropology is pervaded with distinctions, between transcendental and empirical perspectives; between cognition, feeling, and desire; between sensibility and the undboth erstanding. Reinhold’s comprehensive foundationalism results in a “scientific form” with a new “systematic character”, a “thoroughgoing interconnection of . . . material . . . under one principle” (R 109-10, diGiovanni 1985: 84). Where Kant sought a taxonomically exhaustive system laying out conditions of possibility of various different human faculties, Reinhold seeks the unity of all sciences under the single foundational principle of consciousness.

Finally, Reinhold’s emphasis on consciousness as the ground of a whole transcendental system displaced the uniquely human character of Kant’s own transcendental philosophy. For Kant, the world appears as it does because of distinctively human features of cognition, especially the spatial-temporal structure of our sensibility. By contrast, by emphasizing the power of representation, Reinhold sought transcendental conditions of possibility of any world that could be represented by any being, human or otherwise, deriving even space and time not from the particular form of human intuition but from the nature of representation as such.

Reinhold’s attempt to provide a unified, systematic grounding for what he took to be a broadly Kantian transcendental anthropology did not silence Kant’s critics, of course, and in 1792, G.E. Schultze (1761-1833), a little known philosophy professor, wrote a work entitled *Aenesidemus, Or Concerning the Foundations of The Philosophy of the Elements* issued by Prof. Reinhold in Jena, Together with a Defense of Skepticism Against the Pretensions of the Critique of Pure Reason. The work defended a broadly Humean skepticism and rejected Reinhold’s philosophy and the foundationalism that had become the dominant

¹⁵² Reinhold even takes this unifying further, suggesting that his *Philosophy of Elements* will ground both sensible and supersensible objects: “the Philosophy of the Elements must lay down the principles not just for the metaphysics of nature, but for general metaphysics as well – qua metaphysics of knowable objects, but no less qua metaphysics of those supra-sensible objects that are present only to thought” (Foundation 121, diGiovanni 1985:89).

interpretation of Kant's own transcendental anthropology. Its impact was immediate and profound, marking the end of Reinhold's dominance as preeminent defender of Kant. While Reinhold himself quickly faded from the scene after the publication of this criticism of his views, however, Aenesidemus did more to promote than to undermine the foundationalist, reductionist, and systematic direction in which Reinhold took Kant.

In his Review of Aenesidemus (1794), J. G. Fichte (1762-1814) simultaneously named Reinhold as the primary defender of Kantian philosophy and usurped that role.¹⁵³ For Fichte, Schultze's skepticism showed "the precariousness of the position where [philosophical reason] has for the moment come to rest" (diGiovanni 1985: 137). Aenesidemus "concedes . . . that philosophy has so far lacked a supreme, universally valid principle, and that it will be able to elevate itself to the rank of a science only upon the establishment of some such principle" (diGiovanni 1985:138, emphasis added). But Fichte argues, against both Reinhold and Schultze, that this principle need not be based on the power of representation. Fichte thus accepts Reinhold's foundationalism while denying his specific foundation. Moreover, unlike Kant, for whom the attempt to find a common root to the diverse powers of the soul is mistaken and ultimately futile, Fichte seeks "yet a higher concept than that of representation" (diGiovanni 1985: 138). The general reductionist trajectory begun by Reinhold is extended by Fichte and, in varying ways, becomes the guiding thread of German Idealism, which would dominate the appropriation of Kant for the next 30 years.

The key to Fichte's shift away from Reinhold is his move from a "fact [Tatsache] of consciousness" to an original act [Tathandlung], to start not from some dogmatic assertion but from a "highest act," which Fichte identifies with an "act of self-positing" (Fichte 1993:126). As Fichte explains, "the heart of my system is the proposition, 'The I simply posits itself'" (Fichte 1993:398).¹⁵⁴ The self-positing of the I, a self-positing of which one becomes aware only through an "intellectual intuition" of one's own activity, precedes any distinction between subject and object, or between oneself and others. This importantly differentiates Fichte's foundationalism from previous foundationalisms, including Reinhold's. Whereas Reinhold followed Descartes and others in seeking a substantive foundation, Fichte's grounding act provides a "practical"

¹⁵³ Like Reinhold and the German Romantics, Fichte takes himself to be adhering to the spirit of Kant – "my system is none other than the Kantian system" – even if not the letter – "it proceeds in a manner that is entirely independent of Kant's presentation" (Fichte 1994: 4).

¹⁵⁴ See Beiser 2002 for discussion of the ongoing scholarly debate about whether this I should be understood subjectively (as an individual I) or as an absolute I distinct from any individual.

foundation for his philosophy. Fichte takes self-positing as a starting point and shows how it requires positing a “not-I.” In later works, Fichte takes the requirement to posit not-I further, arguing that positing oneself as a free being requires positing a not-I that can issue an invitation (Aufforderung) to activity, and to play this role, the not-I must be free and rational. Thus the I’s self-positing requires positing other free and rational beings, and from this deduction of intersubjectivity, Fichte derives a whole moral and political theory (see his 1796-7 *Foundations of Natural Right* and 1798 *System of Ethics*). An indubitable foundation in the self-positing of the I provides the basis for a systematic philosophy that not only responds to skeptical concerns regarding theoretical knowledge but also grounds a detailed practical philosophy.

Some further points are worth noting in connection with Fichte’s relation to Kant’s transcendental anthropology. First, like Kant, Fichte emphasizes human freedom. An early letter shows Fichte’s admiration for Kant – and rejection of Spinozism – springing precisely from this emphasis on freedom:

The influence that this [Kant’s] philosophy, especially its moral part (though this is unintelligible apart from the Critique of Pure Reason), exercises upon one’s entire way of thinking is unbelievable – as is the revolution that has occurred in my own way of thinking in particular. . . I now believe wholeheartedly in human freedom and realize full well that duty, virtue, and morality are all possible only if freedom is presupposed. I realized this truth very well before . . . but I felt that the entire sequence of my inferences forced me to reject morality . . . If I have time and leisure, I will devote them . . . entirely to the Kantian philosophy. (Fichte 1993: 360).

The “I” that posits itself is free, and its positing is free self-positing. Even the restraints imposed upon the I by the not-I and other free beings must be seen, from the perspective of the I itself, as self-imposed restraints: “the not-I which is supposed to determine the intellect is itself determined by the I – though in this case the I would be considered not as the representing I, but rather as an I which possesses absolute causality” (Fichte 1993: 134). However, Fichte’s admiration for Kantian freedom does not extend to the metaphysical framework that, for Kant, made such freedom possible. Fichte vehemently rejects the doctrine of the thing-in-itself, describing it as “a pure invention which possesses no reality whatsoever” (Fichte 1994: 13), a “piece of whimsy, a pipe dream, a nonthought” (Fichte 1993: 71). Rather than taking “representation” to be primitive and getting trapped by Jacobi’s (and Pistorius’) Affection Problem, Fichte takes Idealism to its limit and denies the need to think of

anything beyond what is posited by an Absolute I (which itself is merely something self-positing, rather than some existent but unknowable thing). Finally, Fichte – like Reinhold – shifts from a specifically anthropological idealism (which argues from the nature of human cognition to a metaphysics of human experience) to an “absolute” idealism that abstracts from anything distinctively human. Whether one interprets Fichte’s “I” as the I of a single individual or an absolute I that stands apart from any particular thinker, the starting point of Fichte’s philosophy is not distinctively human. The “I = I” that marks the beginning of the *Wissenschaftslehre* is, even more than Reinhold’s “principle of consciousness,” a universal, non-anthropological starting point.

By 1800, Kant no longer had control over his own philosophy. His pragmatic and empirical anthropology were largely ignored, and his transcendental anthropology was subjected to vehement criticisms from empiricists, skeptics, and even rationalists. More importantly, Kant’s “defenders” largely left Kant behind, and the world looked not to Kant, but to Reinhold and then Fichte for the comprehensive vision that could resolve the conflicts Jacobi had highlighted between a rational understanding of the world and the freedom necessary for moral responsibility. German Idealism, as it emerged through the work of Reinhold and Fichte, would replace Kant’s anthropological philosophy (with its sensitivity to the diversity of human faculties, the irreducible perspectives that human beings take on the world, and the inherent limitations of these perspectives), with foundationalist, systematic philosophies that sought to find a single absolute and fundamental fact, act, or standpoint from which everything important about the world could be completely understood.

II. Early German Romanticism

The foundationalist, systematic approach of Reinhold and Fichte was not universally accepted as the best way to carry on the “spirit” of Kant’s philosophy. Especially in the context of Hamann and Herder’s critiques of supposedly purely-rational system-building, another group of thinkers – so-called German Romantics – promoted an appropriation of Kant that deemphasized reason and system. Like Reinhold (and Fichte), these thinkers saw themselves as following Kant, but German Romantics focused on a very different Kant. Three central themes characterize the

reception of Kant's anthropology within Early German Romanticism: an attempt to bridge the gap between nature and freedom (and therefore between duty and inclination), an interest in individuality rather than universality, and a focus on the implications of the inscrutable thing-in-itself for the limits of philosophy.

The Romantic interest in bridging the divide between duty and inclination that Kant seemed to assume in his moral philosophy begins with Kant's own insistence in his Critique of Judgment and pragmatic anthropology that there must be a way of bringing together nature and freedom, what we will do and what we ought to do. Drawing on Kant's claim in the Critique of Judgment that the feeling for the beautiful provides a link between nature and freedom, many Romantics emphasized the formative function of art in bridging what Kant's anthropology had seemed to present as an intractable divide. The most important early attempt is found in Friedrich Schiller (1759-1805), who famously offers this witty poetic criticism of Kant's divide between duty and inclination:

Scruples of Conscience

I gladly serve my friends, yet I do it by inclination
And so it often vexes me, that I am not virtuous

Decision

There is no other advice, you must seek to despise them,
And thereby do with abhorrence what duty bids you. (Schiller and Goethe 1796: poems 388-9)

But Schiller's more important contribution to the Romantic reception of Kant's conception of human nature came through his *On the Aesthetic Education of Man*, where Schiller seeks, like Reinhold, to develop a philosophy that is true to the "spirit of the Kantian system" even if not to the "letter of it" (Schiller 1967: 87n2).¹⁵⁵ There Schiller argues,

If we are to be able to count on man's moral behavior . . . as we do on natural effects, it will itself have to be nature, and he will have to be led by his very impulses to the kind of conduct which is bound to proceed from a moral character. But the will of man stands completely free between duty and inclination, and no physical compulsion can, or should, encroach on this sovereign right of his personality. If, then, man is to retain his power of choice and yet, at

¹⁵⁵ With respect to Kant's anthropology in general, Schiller is actually remarkably sensitive to the spirit of Kant's whole anthropology (not merely the transcendental part of it), following Kant in insisting upon distinctions between different powers of soul, developing a philosophy of history with marked parallels to Kant's own, emphasizing the importance of pragmatic-moral anthropology, and even laying out some similar details (e.g. an emphasis on aesthetic illusion) to Kant's own.

the same time, be a reliable link in the chain of causality, this can only be brought about through both these motive forces, inclination and duty, producing completely identical results in the world of phenomena; through the content of his volition remaining the same whatever the difference in form; that is to say, through impulse being sufficiently in harmony with reason to qualify as universal legislator. (Schiller 1967: 17).

Schiller uses Kant's fundamental claim about human beings – that we are both objects in an empirical world and free agents – to argue for a new kind of philosophical task. If we are both free and empirical, then virtue requires not only a strong will that can choose in accordance with the moral law but a set of empirical incentives that will give rise to moral actions as the expression of that will. For Kant, cognition of the moral law appears in the empirical world as a predisposition to personality by which the moral law itself overpowers “inclinations” rooted in the lower faculty of desire. But Schiller's essay “On Grace and Dignity” (1792), objected that “in the moral philosophy of Kant, the idea of duty is proposed with a harshness enough to ruffle the Graces, and one which could easily tempt a feeble mind to seek for moral perfection in the somber paths of an ascetic and monastic life” (Schiller 2006:148) By contrast, Schiller aims for an ethics that would recognize and embrace the whole of human nature: “By the fact that nature has made of him a being at once both reasonable and sensuous, that is to say, a human being, it has prescribed to him the obligation not to separate that which she has united” (Schiller 2006: 147). For Schiller, the “will” that acts in accordance with the moral law has no direct empirical correlate, so bringing together nature and freedom requires developing inclinations that will give rise to the actions duty requires. Thus Schiller sees the unification of reason and sensibility, duty and inclination, as necessary to bring Kant's anthropology to completion. Violating the “letter” of Kant's moral theory by insisting that dutiful actions can and even should be motivated by inclination (empirically) is necessary, Schiller thinks, to preserve the spirit of Kant's anthropological emphasis on both freedom and nature. Even before Kant published his *Anthropology from a Pragmatic Point of View* (1798), Schiller's *On the Aesthetic Education of Man* (1794) puts at center stage what would become Kant's pragmatic anthropology, especially the issue of how to empirically effect moral reform in human beings. Bridging the gap between nature and freedom is an important part of the Romantic appropriation of Kant.

A second key theme within the Romantic appropriation of Kant is the emphasis on human individuality. When Novalis (1772-1802) insists that “the highest task of education is – to seize the mastery of one's

transcendental self – to be at the same time the Self of one’s Self” (Novalis 1797: 207), he does not have in mind Kant’s sort of autonomy as governing oneself by means of a universal moral law, but instead an embracing of “my own, true, innermost life” (Novalis 1797:206) which differs wildly from one person to another. Friedrich Schlegel (1772-1829) praises the one “who has his own religion, his own original way of looking at infinity” (Schlegel 1800:261) and objects that “the Kantians’ conception of duty relates to . . . one’s calling . . . as the dried plant to the fresh flower” (Schlegel 1800:262). For him, “Individuality is precisely what is original and eternal in man; personality doesn’t matter so much. To pursue the cultivation and development of this individuality as one’s highest calling would be a godlike egoism” (Schlegel 1800:264). Schleiermacher (1768-1834) – whose early years were spent in the company of the German Romantics – adds, “What would be the point of the uniform repetition of a highest ideal whereby human beings . . . are all actually identical, the same formula [e.g., the categorical imperative] being merely combined with other coefficients?” (Schleiermacher 2006: 39). True humanity and true freedom are found not in submission to any universal law, even a supposedly self-legislated one, but in embracing and cultivating one’s uniqueness.

Finally, a third key theme in the Romantic appropriation of Kant was an emphasis on the unknowability of ultimate reality. Drawing from Kant’s own insistence that the human mind constructs a comprehensible reality but lacks cognitive access to things-in-themselves, Schlegel rhetorically asks, “Isn’t this entire world constructed by the understanding out of incomprehensibility or chaos?” (Bernstein 2003: 305). But unlike Kant, Romantics do not tidily set chaos to one side and pursue what can be known. Schlegel rejects any philosophy that “inhibits a leap into the suprasensory regions” (Bernstein 2003: 251) and thereby also rejects any philosophy that purports to be complete, systematic, comprehensible, or even consistent: “Philosophy is the real homeland of irony, which one would like to define as logical beauty . . . Irony is the form of paradox (Bernstein 2003: 241). And, put most beautifully because most ironically, “It’s equally fatal for the mind to have a system and to have none. It will simply have to decide to combine the two” (Bernstein 2003: 247).

For many early Romantics, the emphasis on the incomprehensible was linked with a rejection of the Fichtean pretension that the I – and especially some primitive act of self-positing – can be original. Even if self-positing first brings order to experience, Romantics refuse to reject the Kantian notion that something precedes – not temporally but logically or even ontologically – this order. Schlegel puts this in terms of

incomprehensibility or chaos. Hölderlin (1770-1843) describes it as “Being,” a sort of primordial unity out of which subject and object emerge (a unity that cannot therefore really be thought of as a unity *of* subject and object, since this unity *of* already assumes the distinction that Being precedes). Throughout, Romantics shift towards thinking about the world that underlies the emergence of the I, not just the way that the I comes back to the world. In an aphorism that sums up the anthropology of many early Romantics, Schlegel writes, “The human being is Nature creatively looking back at itself” (Schlegel 1800: 261). On the one hand, this aphorism points out the distinctive importance of human beings; we are the way the nature looks back at itself, and we are creative. On the other hand, though, this aphorism shows a way in which Nature is privileged. Rather than seeing Nature as primarily constructed by human beings (as in Kant’s transcendental anthropology), Schlegel sees human creativity itself as the expression of a spiritualized and creative Nature.

Permeating all three developments was an emphasis within early German Romanticism on the importance of art and of feeling.¹⁵⁶ Schiller finds the bridge between nature and freedom in a purely aesthetic, free activity. Schlegel insists that true art expresses the unique inner character of the genius who creates it: “An artist is someone who carries his center within himself” (Schlegel 1800: 263). And Hölderlin points to art – notably tragedy – as revelatory of “the hidden ground of everything in nature” (Bernstein 2003: 193). Precisely because art is not reducible to the categories with which human beings structure the world into a coherent whole, it is capable of mirroring the incomprehensibility of the world: “a classical text [i.e., a poetic creation] must never be entirely comprehensible” (Bernstein 2003: 239). Art represents the world from a perspective, but a perspective that, precisely because it is not universal-conceptual, leaves the incomprehensibility of the world open to further interpretation. Art – including tragedy, poetry, music, etc – provides glimpses of that Being that precedes all categories.

III. Late German Idealism

Fichte marks a high point of Early German Idealism, the pinnacle of foundationalist and I-centered approaches to understanding humans’ place in the world. For Fichte, philosophy was capable of achieving a

¹⁵⁶ In fact, Kant’s *Critique of Judgment* – as a transcendental anthropology of feeling that finds in beauty and the sublime a unity between nature and freedom, a sort of individual freedom in the activity of the genius, and a way of “feeling” the supersensible – could be seen as an early Romantic text, and early Romantics drew heavily from that Critique.

scientific, systematic status through seeing all knowledge as following from the conditions of self-positing of the “I.” Early German Romantics resisted Fichte’s subjectivism,¹⁵⁷ systematicity, and rejection of limits for human knowledge. The next major stage of German Idealism was dominated by two figures – Friedrich Schelling¹⁵⁸ (1775-1854) and G. W. F Hegel (1770-1831) – who were closely tied to German Romanticism. Schelling, Hegel, and Hölderlin lived and studied together in Tübingen, and all three worked together on an “Oldest System-Programme of German Idealism,” which lays out a sort of program for “a new mythology . . . of reason.” But while Schelling and Hegel shared some central Romantic convictions, they fundamentally accepted the need for systematic philosophy.

Among the most important problems facing the later German Idealists was the problem of how to reconcile the empiricist and especially Romantic recognition that humans’ interpretive activity is something that emerges out of a pre-existing Nature or Being with the insights of Kant’s properly transcendental anthropology. Two issues here were particularly important. The first was how to preserve cognitive access to the world. Unlike the Romantics, Schelling and especially Hegel refused to accept the absolute incomprehensibility of ultimate reality. Like Reinhold and Fichte, they sought some way of avoiding commitment to an absolutely unknowable thing-in-itself. But this required responding to Kant’s arguments that the Copernican turn that makes knowledge possible also limits the scope of that knowledge. The second issue was how to preserve the from-the-inside dimension of transcendental anthropology while integrating human beings into a more comprehensive natural whole.

Early Idealism and Romanticism provided some resources for systematically integrating human beings into an absolute whole. Fichte’s “Absolute I” provided one avenue. In Schelling’s hands, the “Absolute I” was transformed into merely “the Absolute” as such. Hölderlin had already suggested that “Being expresses the combination of subject of object” and specifically directed this conception of Being against Fichte’s “I=I” in which subject and object are united through combining what is

¹⁵⁷ See Beiser 2002: 217ff. for discussion of the scholarly debate about whether the ascription of subjectivism to Fichte is justified.

¹⁵⁸ For the purposes of this brief summary, I focus on the early work of Schelling, ending in the early 1800s. Schelling is famous for constantly shifting philosophical views, but here I aim for a more-or-less unified view of his early works. For discussions of Schelling’s whole corpus, see Bowie 1993, Snow 1996, and White 1983.

already posited as separate. Schelling follows Hölderlin in arguing that Fichte's "I" must be conceived of as something that emerges from a more basic "Absolute." Whereas Fichte and Reinhold saw Kant's philosophy of freedom as a way to avoid a Spinozism that he saw as inimical to morality and religion, Schelling made use of this same emphasis on freedom to redeem Spinozism by showing how Being itself can ground freedom. Schelling's project has three parts, which we might see as the stages towards fusing Fichte with Spinoza and thereby integrating a transcendental and normative perspective into a Romantic conception of human beings as emerging from Nature.

First, in his *Naturphilosophie* (Nature-Philosophy) of 1797, Schelling sets out to transform Spinoza by developing what we might think of as a transcendental cosmology. Rather than looking at the world of appearance from the perspective of human experiencers, as in Kant's transcendental anthropology, Schelling attempts to think of the world of appearances from the perspective of nature itself. Schelling thus seeks to explain the world out of which human knowers could have emerged, rather than simply the world known by human beings. Here the question is not "what are the conditions in human nature that make it possible to experience a world?" but "what are the conditions in Nature that allow it to be the sort of thing that can be experienced?" Like Spinoza, Schelling adopts a God-eye view on the world, seeing it as a whole that manifests itself in different ways. But for Schelling, nature is not a set of brute mechanical interactions. Drawing from the spirit (though certainly not the letter!) of Kant's *Critique of Judgment*, Schelling sees nature as purposive, an "original productivity" (Bowie 2003:75). And just as Fichte argued that the I must limit itself in order to become conscious of itself, Schelling argues that Nature must limit – or "inhibit" – its own productivity in order to emerge from the realm of vital forces to appear as an "object" for a human perceiver. In thinking of Nature itself as an original productivity, Schelling transcends, from the side of nature, the determinism that seemed to rule out freedom (and with it morality). Whereas Fichte (and Kant) thought freedom could be preserved only by prioritizing human activity, Schelling points out that Nature itself can be conceived as a set of active forces. Rather than protecting human freedom from Nature, Schelling argues that human freedom emerges from freedom already in nature. In contrast both to traditional compatibilisms (which change our concept of freedom so that it fits with natural necessity) and to Kant (who posits nature and freedom in separate realms), Schelling changes our concept of nature so that it fits with our conception of freedom as something fundamentally active.

Second, in his *System of Transcendental Idealism* of 1800, Schelling considers the problem of the relationship between nature and freedom from a Fichtean (rather than Spinozist) starting point.

[I]n the fact of my knowing, objective and subjective are so united that one cannot say which of the two has priority . . . [so] there are two possibilities.

A. *Either the objective is made primary, and the question is: how a subjective is annexed thereto[?] . . .*

B. *[Or] the subjective is made primary, and the problem is: how an objective supervenes, which coincides with it?* (Schelling 1978:6-7)

Naturphilosophie addressed the first possibility, concluding that Nature itself must have properties associated with subjectivity, in particular a sort of spontaneity that comes to emerge in consciousness. *System* addresses the second possibility. Starting from the perspective of the knowing I, Schelling considers what is necessary in order for that I to have knowledge of a world that is not itself a subjectivity. And his central claim in *System* is the converse of the central claim in *Naturphilosophie*. Just as Nature must be united with subjectivity and thus must be spontaneous, so the free, spontaneous, and *conscious* “I” must be united with Nature and thus grounded in unconscious elements. Schelling thus posits “an identity of the nonconscious activity that brought forth nature and the conscious activity expressed in willing” and exhibits “this simultaneously conscious and nonconscious activity . . . *in consciousness itself*” (Schelling 1978:12). In the process, Schelling draws from Fichte’s argument that consciousness depends upon the not-I and echoes Holderlin’s appeal to a Being more primordial than the subject-object dichotomy to argue for a sort of incoherence in the very categories in terms of which Kant cast his famous “Copernican turn”:

[T]he distinction between what comes from without and what comes from within . . . stands in need of justification . . . But . . . I posit a region of consciousness [really, *unconsciousness*] where this separation *does not exist*, and where inner and outer worlds are conceived as interfused. (Schelling 1978:74)

Like his *Naturphilosophie*, *System* aims to undermine the Kantian distinction between nature and freedom, now by showing that freedom itself has an unconscious, “natural,” dimension. And for Schelling – arguably following the lead provided by Kant in his *Critique of Judgment*, but certainly following the lead of Romantics like Schiller – the unconscious dimension of human subjectivity has important

implications for the sort of “knowledge” that one can have of self and nature. Since *conscious* knowing is only partial and derivative, access to nature as well as oneself can only take place most fully in the context of a process of awareness that combined conscious and unconscious elements. This awareness takes place in art, in both “the artist . . . lost in his work” (Schelling 1978:75) and those aesthetically reflecting on the work of art:

The work of art reflects to me what is otherwise not reflected by anything, namely that absolutely identical which has already divided itself even in the self . . . [A]rt is at once the only true and eternal organ and document of philosophy, which . . . speak[s] to us of what philosophy cannot . . . , name the unconscious element in acting and producing, and its original identity with the conscious. (Schelling 1978:230-1)

One recent commentator has gone so far as to claim, “The *System* can be described simply as an elaborate proof of the proposition that only aesthetic intuition can provide, in the production of the work of art, the reconciliation *in consciousness* of finite and infinite, conscious and unconscious activities, without assimilating the unconscious to the conscious” (Snow 1996: 120). Art is the way in which the primordial unity of subject and object, of nature and freedom, of unconscious and conscious, becomes manifest *to consciousness*. Art as the intuitive awareness of original unity is Schelling’s Romantic conclusion to Fichte’s reflection on the subjective conditions of knowing.

Finally, Schelling pulls together these elements into a single “Identity Philosophy” that emphasizes that the unity of subject and object in knowing and doing is always already rooted in a primordial “Absolute Identity” that precedes the separation into subject and object. In his later work, Schelling moves away from the sympathetic approach to Fichte with which his 1800 *System of Transcendental Idealism* began; rather than saying that “the result of the operation is bound to be the same” whether one starts with Nature or with the I, Schelling’s later emphasis on the Absolute involves a fundamental subordination of the “I” to a whole (*das Alls*) that precedes it:

Since Descartes, the *I think, I am*, has been the basic mistake of all knowledge; thinking is not my thinking, and being is not my being, for everything is only of God or the whole. (*Aphorisms as an Introduction to the Philosophy of Nature* (1804))

It is not the I that knows, but the whole [universe] through me . . . [Without] some knowledge totally independent of all subjectivity . . . of . . . the simple One . . . , we would be eternally locked inside the sphere of subjectivity. (*System of the Whole of Philosophy and of Nature in Particular* (1804), I/6:140, 143)

With this shift away from subject-centered idealism, however, Schelling also solidifies his criticisms of attempts to come to complete knowledge – of self, Nature, or the Absolute – through concept-based philosophy; art and intuition, rather than conceptual analysis or objective knowledge, become the access points for contact with the Absolute, and no “understanding” of that Absolute can ever be complete.

In the end,¹⁵⁹ Schelling shares with early German Idealism a commitment to overcoming the dualisms in Kant’s distinction between appearances and things-in-themselves, and in affirming a Kantian emphasis on combining a philosophy of human freedom with a natural philosophy that could do justice to the empirical world. But Schelling adopts from the Romantics a skepticism about ultimate conceptual knowledge of the world, an emphasis on art and creativity, and even – from Hölderlin in particular – an insistence that the really real must *precede* the distinctions between subject and object, nature and freedom, that are so central not only to Kant but also to Reinhold and Fichte.

The climax of German Idealism comes with G.W.F. Hegel. Hegel got his philosophical start through working with Schelling in Jena during a period in which Schelling had just taken the philosophical mantle from Fichte. By the time Hegel surpassed Schelling to become the dominant voice in German philosophy, Kant’s role in German philosophy had become largely symbolic. The particular ideas associated with Kant were those appropriated or criticized in Fichte, Schelling, and the Romantics, and Hegel interacts primarily with these later figures. Nonetheless, Hegel pulls together many of the central criticisms of Kant’s philosophy that emerged in the decades after the publication of the first Critique. Like Schelling, he seeks to appropriate but move beyond not only the Early German Idealists but also the German Romantics. Unlike Schelling,

¹⁵⁹ Well, not quite the end. I have cut short my discussion of Schelling here at his 1804 *System*. His thought continues to undergo significant change for 30 more years. *Of Human Freedom*, his last published work (in 1809) is particularly valuable its challenge to Kant’s conception of freedom, but it is also a work that arguably takes Schelling beyond German Idealism, and the interpretation of the work is widely contested (compare, e.g., White 1983, Snow 1996, and Heidegger’s very important treatment in Heidegger 1985). As important as the work is, getting into its details would take the present chapter too far astray.

Hegel's sympathies lie more with the system-building Idealists than the aesthetic Romantics. Thus where Hegel follows earlier thinkers in rejecting Kant's distinction between appearances and things-in-themselves, Hegel does so – unlike Schelling or Hölderlin – in the service of Absolute *Knowledge*. In this Introduction to his *Phenomenology of Spirit*, after laying out various ways in which a “mistrust of Science” might come about through thinking about the gap between our faculties of cognition and the Absolute that “Science” aims to know, Hegel says, “if . . . Science, . . . in the absence of such scruples [about its possibility] gets on with the work itself and actually cognizes something, it is hard to see why we should not turn around and mistrust this very mistrust” (Hegel 1979: 47). Hegel aims to undermine both Kant's idealism about knowing things in themselves and Schelling's skepticism of *scientific* knowledge of the Absolute by actually establishing that knowledge.

In order to do this, Hegel develops his own philosophical methodology, a “logic” distinct from Kant's “transcendental” style of argument, Fichte's foundationalism, Romantic aphoristic and mythological philosophy, and Schelling's appeal to (aesthetic or intellectual) intuition. In some respects, this dialectical logic can be seen as a systematizing the reasoning Kant used in his Antinomies in *The Critique of Pure Reason*. In the third antinomy, for example, Kant shows how one can prove both determinism and the necessity of a freedom that contradicts determinism, and this “antinomy” – proof of both sides – leads him to diagnose and reject the underlying premise of both sides: transcendental realism. In a similar way, Hegel's dialectical logic involves moving through positions that are revealed as “one-sided,” in such a way that one first moves from, say, an overly subjective approach, to an overly objective approach, and then moves forward by rejecting the common one-sided assumption underlying both approaches. For example, Hegel's *Phenomenology* starts with the naïve view that what is real is simply “This,” that which is immediately present and certain to the senses, but shows that this “This” ends up being empty without some conceptual determination. This challenge to “Sense-Certainty” does not lead to an empty skepticism, but to a further position, in which one privileges abstract concepts over immediate sense experience. But this view has problems of its own, and that leads to a further refinement, and so on. In his moral philosophy, Hegel begins with the view that freedom is simply doing what one feels like doing (the “natural will,” see Hegel 1991: §11), but this does not allow one to rise above contingently-given inclinations, so one pursues a “negative will” (Hegel 1991: §§5, 15), which ends up empty of content, and so on. The general strategy here is to start from some naïve given conception of reality and our knowledge of it and

“negate” this conception by showing its immanent conflict with itself, but to negate it in a way that “determinately” gives rise to a more sophisticated conception of reality (see especially Hegel 1979: §§77ff.).

Hegel’s dialectical arguments lead to a more holistic conception of human beings than Kant’s. Hegel not only objects to distinguishing empirical objects from things-in-themselves, but also – like Schelling – incorporates various Kantian dichotomies into a single, coherent conception of the self (and world) that combines objective and subjective, volition and cognition, understanding and sensibility. Hegel accepts the force of the Affection Problem and rejects the unknowable “thing-in-itself.” He appropriates and broadens Schiller’s critique of Kant’s separation between duty and inclination, insisting that “what is actual becomes rational and what is rational becomes actual” (cited in Hegel 1991: viii, cf. § 1). For human motivation in particular, “it is an empty assertion of the abstract understanding to require that only a [moral] end shall appear [autonomously] willed . . . , and likewise to take the view that, in volition, objective and subjective ends are mutually exclusive” (Hegel 1991: 151). Like Romantics, Hegel seeks unity between subjective inclinations and moral duties. (He finds this, first, in love, but eventually also in the feelings of honor, trust, and patriotism that lead one to do one’s duties in society and the state.) Like Kant, Hegel insists that human particularity must be subordinated to universal principles, but his conception of “universality” is concrete and social rather than abstract. Hegel argues that Kant’s conception of the categorical imperative as an “abstract universality” makes morality an “empty formalism.” Without “material from outside,” the categorical imperative gives no particular duties, since *anything* can be universalized if one does not care about any particular consequences: “the fact that no property is present is in itself no more contradictory than is the non-existence of this or that individual people . . . or the complete *absence of human life*” (Hegel 1991: 162). But while Hegel considers Kant’s abstract ethics too one-sided, he sees the Romantic emphasis on feeling and individual creativity as equally one-sided and empty (see especially Hegel 1991: §140-41). Instead, Hegel proposes that “ethics” – as opposed to an abstract Kantian “morality” – must be situated in particular social and political contexts. He approvingly quotes “a Pythagorean” who advised a father asking about how best to “educat[e] his son in ethical matters . . . : ‘Make him a *citizen of a state with good laws.*’” (*Philosophy of Right*, § 153, Wood, p. 196). Hegel develops a complicated description of ethical life in which moral duties are given by a person’s place in a familial, economic, and political order.

Hegel thinks that humans can get insight into the legitimacy of the existing world – including one’s social order – because insight emerges from the unfolding of the world itself. Like Schelling, Hegel sees art (and religion) as important sources of insight into the rationality of the world, but against Schelling, he argues that *ultimate* insight into the world’s rationality is provided by philosophy. Moreover, with Herder and Hamann, Hegel insists that human beings must be understood historically and, with Schelling, he sees human rationality as an historical achievement rather than an eternal condition or a transcendental fact. But Hegel insists that this historical development is rational and rationally comprehensible, that world history *justifies* its stages. Human beings arise out of nature through an inevitable process of Spirit, or the Absolute, coming to know itself. And for Hegel, ethical and epistemic norms not only arise out of the concrete historical situations in which one finds oneself but are, at least in principle, ultimately conducive to Absolute Knowledge. Hegel ends up with a conception of humans as agents and knowers situated in terms of a nature, history, and social life that are rationally comprehensible through a Science of dialectical reason. Reason is not the purified faculty to which Hamann objected in his *Metacritique* but an ability made possible through the use of language shaped in particular historical societies. Freedom is not an abstract metaphysical concept, not an “uncaused cause” or noumenal thing-in-itself grounding an empirical world, but an historical and social accomplishment. Human beings are free because (or insofar as) we are self-knowing members of well-ordered communities at the right moment(s) in world history.

IV. Conclusion

The immediate responses to Kant’s anthropology focused on his *transcendental* anthropology. Kant’s pragmatic and empirical anthropology had important influences, but much less important than he had hoped. Kant’s student Herder was influenced at least as much by the proto-pragmatic anthropology he was exposed to in Kant’s lectures as by any transcendental anthropology developed later. Schleiermacher famously – and quite negatively – reviewed Kant’s *Anthropology from a Pragmatic Point of View*. Kant’s general approach to empirical anthropology, and especially his tripartite conception of human mental states, helped set the stage for later developments in empirical psychology. And Kant’s conception of human evil became a major focus of discussion not only by Schelling and Hegel but also by Kierkegaard. But in the end, Kant’s empirical and pragmatic anthropology was largely

marginalized during the 19th century. Herder's more radical historicism and Platner's physiological anthropology, rather than Kant's pragmatic anthropology, drove the agenda of empirical research on human beings throughout 19th century Germany (see Zammito 2002:253).

As we have seen in this chapter, there were a wide range of responses to Kant's transcendental anthropology. There are, of course, important figures during this period that this chapter has passed over. Among these, two of the most important for later developments (especially Nietzsche) are Schopenhauer and Kierkegaard. Schopenhauer stands largely outside of the traditions of German Idealism and Romanticism that characterized the early reception of Kant's philosophy, though like these movements, Schopenhauer saw himself as the true heir – but also corrector – of Kant: “While I start in large measure from what was achieved by the great Kant, serious study of his works has . . . enabled me to discover great errors in them” (Schopenhauer 1966: xv). Like the early German Idealists, especially Reinhold, Schopenhauer sought to develop a more simplified, foundationalist structure to Kant's philosophy. But whereas early Idealists such as Reinhold drew largely from Descartes as a model for foundationalism, Schopenhauer was more sympathetic to Berkeleyan idealism, and his reading of Kant – which focused on the first edition of the Critique of Pure Reason, the version that Feder and Garve referred to as merely a new Berkeleyanism – pushed Kant in this Berkeleyan direction. Schopenhauer's magnum opus, *The World as Will and Representation*, starts with the fundamental claim that “The world is my representation” (Schopenhauer 1966: 3). But Schopenhauer simplifies both the argument for this position and the transcendental structure of cognition itself, ridding Kant's argument of what many saw as unnecessarily complicated arguments and artificial taxonomies of “categories,” “faculties,” and so on. Unlike many early responses to Kant, Schopenhauer does not reject a distinction between appearances (the World as Representation) and something akin to an underlying thing-in-itself, which Schopenhauer calls the Will. Unlike Kant, however, Schopenhauer restricts *all* cognition (not just space and time) to the world as representation, which means that the relationship between Will and Representation cannot, in any sense, be a relation of causation or grounding. Schopenhauer aims to get around the Affection Problem by positing that there is a *single* world that is known by human subjects *as* a world of representation, but of which human beings have direct “consciousness” as Will in awareness of their own wills (Schopenhauer 1966: 99-100). Given that all categories are precluded from the analysis of the Will, however, Schopenhauer argues that “the will, considered as such . . . , knows no plurality and consequently is *one*”

(Schopenhauer 1966: 128). With that move, Schopenhauer moves in quasi-Kantian way toward the undifferentiated “Being” that had been posited by Hölderlin. From there, Schopenhauer develops his own version of the developmental story that we saw in Schelling and Hegel. The Will is engaged in a constant struggle against itself, emerging in higher and higher “Forms,” culminating (so far) in what we think of as individual human wills. Since this whole activity of the Will is struggle against itself, however, it is also a sort of suffering, so Schopenhauer moves on to recommend a sort of asceticism that would turn the Will against itself, ending in a “death . . . which breaks up the phenomenon of this will, the essence of [which has] long since expired through free denial . . . ; for him who ends thus, the world has at the same time ended” (Schopenhauer 1966: 382).¹⁶⁰ A few decades later, Nietzsche would adopt much of this picture of the Will struggling against itself in a constant path of suffering and self-overcoming, but reject the ascetic pessimism and, embracing suffering, also embrace higher and higher manifestations of Will.

Where Schopenhauer represents an alternative sort of system-building appropriation of Kant, Kierkegaard inherits the anti-systematic interests of the Romantics (whom he read sympathetically) and the late Schelling (whose lectures he attended). Kierkegaard’s thought cannot be easily summarized – one of his pseudonyms “dreads what is . . . frightful, that one or another enterprising summarizer [and] paragraph-gobbler . . . will cut him up into paragraphs” (Kierkegaard 2006:6) – but one important trend of this thought is an emphasis on what he calls “Subjective Truth,” or, more radically, the notion that “Truth is Subjectivity” (Kierkegaard 1992: 189ff.). Briefly, one might understand this as a sort of fruit of taking Kant’s Copernican turn seriously while rejecting the principled distinction between cognition and volition. The point, then, is that what can be taken to be “true” by an individual is constrained not merely by forms of cognition but also by forms of volition, or *relevance*. This leads to an emphasis on the importance of truths to which one can relate *passionately*, and eventually to an anti-systematic conception of Christian faith any ethical or epistemic universals (see, e.g., Kierkegaard 1985, 2006).

Even with the addition of these figures, there are some general aspects of the reception of Kant that are common amongst most of his immediate successors. Many of Kant’s ideas were retained, such as his

¹⁶⁰ This account skips several important steps in Schopenhauer’s argument, and his entire (important) discussion of art (especially music) as a will-less knowing of the Will as such, but these aspects of his view would go beyond the needs of the present chapter.

emphases on human freedom and some version of his Copernican turn, but most of Kant's successors sought to move away from Kant's sharp distinction between appearances we can know and "things-in-themselves" we cannot. There was also a general effort to overcome the wide variety of distinctions and dichotomies within Kant's system in favor of a more holistic approach to human beings, whether in the form of Reinhold's reemphasis on representation as such, or Herder's (and others') rejection of the sharp division between cognition and volition, or Schelling and Hegel's rejection of the subject-object dichotomy, or Schiller and Hegel's efforts to overcome the distinction between inclinations and duty. Throughout the period of Kant's immediate reception, there was a tendency to push Kant's philosophy as a whole towards extremes of systematicity, either making it much more systematic (Reinhold, Fichte, Hegel) or resisting systematic knowledge altogether (Herder, Romantics, and to some extent Schelling). Kant's interest in human history was also increasingly transformed and radicalized during this period towards something like a transcendental historicism that sees not only particular human tendencies and structures as historically developing but even the basic structure of human reality as subject to historical change.

Chapter 7: 19th Century Alternatives—Darwin, Freud, Marx, and Nietzsche

At the time of his death in 1831, Hegel was the pre-eminent philosopher in Germany (and arguably the world) and philosophy, at least in Germany, was the most important discipline in the academy. Within a decade, the philosophical world in Germany had largely moved beyond Hegel, but without an obvious replacement. Schelling, who held Hegel's old position at the University of Berlin, was an able critic but failed to generate many philosophical followers of his own. For the rest of the 19th century, philosophy in Germany was largely dominated by a variety of neoKantian philosophical movements, attempts to get "back to Kant!" in ways that would be honest to the scientific and historical developments of the 19th century. At the same time, since Hegel's death no single philosopher has dominated the philosophical scene in any part of the world, and philosophy has seen its role increasingly diminish. Hegel's philosophical system – like Kant's – included a detailed logic, an account of the nature of being, an investigation of the basic principles of physics and biology, explanations of political and social organization both historically and in his own day, an account of the proper role of the state, a detailed aesthetics (including, in Hegel's case, a history of art), interpretations of the French Revolution and the rise of the nation-state, and so on. For many of these questions, it is hard to imagine that *philosophers* provide the best answers today, and for none of them is it clear *which* philosopher one should turn to. Many people even suspect that any questions for which one needs a philosopher are likely to be those for which there cannot be any real answer. Whereas many in Hegel's time held out hope that philosophy could provide solid, lasting, and comprehensive answers to all important questions, today the most plausible "Theories of Everything" seem to come from physics, or biology, or even social sciences. This shift is particularly dramatic in terms of the question, What is the human being? For Kant, this question was simply *equivalent* to philosophy as such. Even for Kant's immediate successors, the question was a properly philosophical one. Today, however, most would be more likely to turn to the sciences – biology, psychology, or sociology – for a rigorous answer to the question. And even those who worry that the sciences might miss something essential about human beings are more likely to follow the Romantics into art or literature than to embrace philosophy.

Our present condition did not emerge from a vacuum. While neoKantianism dominated the philosophical scene in Germany after Hegel, many figures arose during the mid-to-late 19th and early 20th

centuries that have shaped the context of our current thinking about human beings in ways that cannot be ignored. And several of the most important such influences relate directly to both Kant's own philosophical anthropology and to the way in which that philosophy was appropriated by Kant's immediate successors. In this chapter, I look at four episodes in the development of this contemporary perspective. In each case, I point out both the ways in which these developments pose significant problems for Kant's conception of human beings and the ways in which they draw from (or mesh with) important dimensions of Kant's thought. I start with the figure who marks, in retrospect, the most direct successor to Hegel: Marx. Marx's historical materialism and his conception of human beings as socially and materially determined beings not only reflected an important response to and appropriation of earlier philosophical accounts of the human being, but it also helped set the stage for the emergence of the social and historical disciplines that play such important roles in our conceptions of human beings today. Second, I turn to the 19th century thinker least closely connected to German philosophy: Darwin. I focus here on two important roles of his theory of evolution by natural selection. This theory solidified the status of biology as a natural science, and it offered a theoretical structure for a new biological account of human beings. Third, I look at the most important purely *philosophical* development in 19th century philosophy: the work of Friedrich Nietzsche, and especially his genealogical revaluations of values. And finally, I look at the development of these trends and the emergence of scientific psychology in the work of Sigmund Freud.

I. Marx and the Rise of Human Sciences

Marx (1818-1883) is one of the dominant intellectual figures of the 19th century, and Marxism was undeniably one of the greatest political influences on the 20th century. Marx is typically thought of more in connection with the revolutionary *Communist Manifesto* that he co-authored with Engels than with his close analysis of Hegel's *Philosophy of Right* or even his own magnum opus, *Capital*. In this brief section, however, my emphasis will be on Marx's philosophical conception of human beings and the way in which this conception affected his communist ideas. Of all the figures discussed in this chapter, Marx most directly follows from the issues and debates discussed in chapter six. Marx was a close reader of Hegel, and from Hegel he took the core idea that human beings are essentially constituted by historically situated social interactions and that human freedom requires embodiment in the world through socially recognized work. But Marx rejects what he saw as

Hegel's "uncritical" political conservatism that teaches a mere "restoration of the existing empirical world," seeking to redeem that world in abstract thought rather than to *change* it (Marx 1963: 201; Marx 1994:83). The result is a conception of human nature as a historical accomplishment that depends upon certain economic conditions, conditions not yet realized under the current (both then and now) capitalist economic system. In this short section, I start with a discussion of Marx's human ideal, then turn to an account of the ways in which this ideal is not yet achieved because of the alienation in capitalist (and pre-capitalist) forms of life, and then end with a discussion of Marx's dialectical materialism and revolutionary, change-oriented approach to philosophy.

For Marx, humans are relational, and our most important relationships are with nature, society, and ourselves. With respect to the natural world, Marx follows Hegel in emphasizing that what it means to be human is in part to make the world in our own image, not merely in the Kantian sense that we see it through human categories, but in the more practical sense that we *work* on nature – whether physically or intellectually or both – to make it into a human world. "The human being . . . begins to distinguish himself from the animal the moment he begins to *produce* his means of subsistence . . . By producing food, the human being indirectly produces his material life itself" (Marx 1970: 42, also in Marx 1994:107). Put another way, "It is just in his work upon the objective world that the human being really proves himself . . . By means of it nature appears as *his* work and reality. The object of labor, therefore, is the *objectification of man's species-life*" (Marx 1963: 128). By working on nature, human beings make real the sort of beings that they are: "In the type of life activity resides the whole character of a species, its species-character" (Marx 1963: 128). Thus in the ideal, humans should "create the human being by human labor" that effectuates "free, conscious activity [which] is the species character of human beings" (Marx 1963: 128, 166). While the species character of animals need involve nothing more than perpetuating the physical and instinctive characteristics of their species, human engagement with the world should bring about expressions of the freedom that is characteristically human. Thus Marx explains that in contrast to animals, "man produces when he is free from physical need and only truly produces in freedom from such need" (Marx 1963: 128), for example, when a human being "constructs . . . in accordance with laws of beauty" (Marx 1963: 128).

The emphasis on human *work* as our proper relationship with nature is combined, in Marx, with a strong emphasis on the *social* nature of human beings. The human being is a "species-being" that "is in his

existence . . . a social being” (Marx 1963: 154). Marx takes as an example his own “scientific work,” since science – as the pursuit of truth about the world – might seem to be something that could be carried on in isolation from others. But Marx insists that this activity is “social, because human” (Marx 1963: 157) in that even “the material of my activity – such as language itself” is a social contribution to that activity, and “what I myself produce I produce for society, and with the consciousness of acting as a social being” (Marx 1963: 158). To be truly human, one must live and work in the context of relationships to society.

Finally, self-realization – our proper relationship with ourselves – takes place through actualizing free, conscious activity of socially-situated work in our own particular way. Marx emphasizes that even one’s individuality matters as the individuality of a social being: “the human being is a unique individual—and it is just his particularity which makes him an individual, a really individual communal being” (Marx 1963: 158). My particular strengths and weaknesses show up and are relevant only within a social context. Just as Marx’s unique political and social theorizing depends upon a social context, so too does the activity of any human individual. For Marx, one truly becomes who one is when one “makes the community (his own as well as those of other things) his object both practically and theoretically, [and] also (and this is simply another expression for the same thing) in the sense that he treats himself as the present, living species, as a *universal* and consequently free being” (Marx 1963: 128). As an individual communal being drawing from society to contribute to society, our labor helps define and expand the parameters of human possibility. One relates to oneself properly, as a “species being,” when one’s individual life becomes a means for expressing the goods of the species in free and creative ways.

For Marx, then, humans are those beings that engage in free, conscious, social activity. Unfortunately, according to Marx, one living in a capitalist society “exists only as a *worker* and not as a human being” (Marx 1963: 137), and, what’s worse, “the worker sinks to the level of a commodity, and to a most miserable commodity” (Marx 1963: 120). The dehumanization of workers takes place through what Marx calls “alienation.” In general, alienation occurs when what should be an essential aspect of one’s own humanity is made foreign – or “alien” – to oneself. Marx diagnoses four key forms of alienation within capitalism: alienation from the products of one’s labor, from other people, from oneself, and from humanity.

Alienation from the products of one’s labor is the simplest to comprehend: “the alienation of the worker in his product means not only

that his labor becomes an object, assumes an *external* existence, but that it exists independently, *outside himself*, and alien to him, and that it stands opposed to him as an autonomous power” (Marx 1963: 122-3). *Any* product, simply by virtue of being a product, will stand independent of its maker in some sense. When Marx develops a scientific theory or when an artist makes a work of art or when a good cook makes a delicious meal, the theory and the work of art and the meal are independent of their creators. But in all of these cases, the work is the concrete *expression* of the free activity of its maker. Insofar as the theory is studied or the work contemplated or the meal eaten, its maker accomplishes his own ends and the activity is merely completed in the external existence of the product. But within capitalist economies, workers no longer produce for the sake of their products but for the sake of *wages*.¹⁶¹ In such a system, the *goal* of the worker is wages that are used for sustaining (animal) life, and the independent product – the automobile or the blue jeans or even the artwork marketed to wealthy collectors or the meal sold to paying customers – is an *alien* product, something that does not express the free and conscious activity of its *human* maker but merely the mechanical operations of a worker. The result, for Marx, is devastating: “The more the worker expends himself in work the more powerful becomes the world of objects which he creates in the face of himself, the poorer he becomes in his inner life, and the less he belongs to himself” (Marx 1963: 121). Rather than taking an alien nature and transforming it in the light of free human activity, the workers in a capitalist economy imbue alien nature with their own labor in order to make it an even more powerful and *more alien* force working against them. One ends up alienated not only from the immediate products of one’s labor, but – because labor is the way in which the *world* becomes one’s own – one is alienated from the “sensuous external world” as a whole (Marx 1963: 125).

Moreover, for Marx, within capitalism, human relationship become increasingly corrupted. People are valued for the extent to which they can generate higher profits. This valuing can be direct and obvious, as when employers pay all workers the same wages based on their productivity, treating them as more-or-less efficient product-generators rather than as human beings in a cooperative enterprise of transforming nature for the good of society. But alienation in human relations happens even in more intimate contexts, where one increasingly regards friends and neighbors, and even parents and children, in terms of how

¹⁶¹ This emphasis on wage-labor is distinctive of capitalism, but alienated labor is not. In early epochs, one might work for the sake of God, for instance, but one’s product would still be alienated, expressing the will of God rather than one’s own free, conscious activity.

much they help or hinder one's economic interests. Think here of parents who make decisions about how much time to spend with children based on what they can afford, or of the ways in which friends seek to ensure that bills are split equally so that no one falls into another's debt. And even more subtly, human relationships end up being crafted around communities of common interest, where "interests" end up being largely a matter of one's situation within the labor market. As Marx puts it, "every man regards other men according to the standards and relationships in which he finds himself placed as a worker" (Marx 1963: 129).

The alienation from the products of one's labor and from truly human relationships with others leads to an alienation from one's *self*. For one thing, alienation from the *products* of labor quickly becomes alienation from the *activity* of labor:

the work is *external* to the worker . . . [I]t is not a part of his nature; and . . . consequently he does not fulfill himself in his work but denies himself, has a feeling of misery rather than well-being, does not develop freely his mental and physical energies but is physically exhausted and mentally debased . . . His work is not the satisfaction of a need, but only a *means* for satisfying other needs . . . This is the relationship of the worker to his own activity as something alien and not belonging to him, activity as suffering (passivity), strength as powerlessness, creation as emasculation, the *personal* physical and mental energy of the worker, his personal life (for what is life but activity?), as an activity which is directed against himself, independent of him and not belonging to him. This is the *self-alienation* as against the . . . alienation of the *thing*. (Marx 1963: 124-6).

What starts as the selling of the products of labor quickly transforms labor itself, one's life-activity as such, into something that is not part of who one *is*, but something to be traded *away* in exchange for satisfying basic needs. But because, for Marx, life *is* activity, who one *is* is what one does and one's place in society, alienation from one's *activity* is alienation from one*self*.

The result of this three-fold alienation from the sensuous world, other people, and oneself, is an alienation from true humanity. "[T]he human becomes animal" as people

find fulfillment not in work and relationships but in food, sex, and at best "decorating and personal adornment" (Marx 1963: 125). Rather than

the activity of the individual contributing to defining the species, one puts one's distinctively human traits – one's language, reason, relationships with nature and others – to the service of merely animal ones:

alienated labor . . . makes *species-life* into a means of individual life . . . For labor, *life activity, productive life*, now appear to man only as *means* for the satisfaction of a need, the need to maintain his physical existence. (Marx 1963: 128)

Even if, say, one constructs a home in accordance with aesthetic ideals, in a capitalist economy one does so only so that one can meet merely animal needs, so that one can get better wages and thereby more food, sex, entertainment, and so on. Capitalism is precisely that condition “in which every natural and social characteristic of the object is *dissolved*, in which private property has lost its natural and social quality (and has thereby lost all political and social disguise and no longer even *appears* to be connected with human relationships) and in which the *same* capital remains the *same* in the most varied natural and social conditions, which have no relevance to its *real* content” (Marx 1963: 139). In the end, “alienated labor . . . alienates from man his own body, external nature, his mental life, and his *human life*” (Marx 1963: 129).

This account of alienation requires Marx to answer to the question “What is the human being?” for *future* humans, those not living under alienating economic conditions. For Marx, humanity is the end of a process – “world history is . . . the creation of the human being” (Marx 1963: 166-7) – and a process that is not yet finished: “the *human level* . . . will be the . . . *future* of [currently capitalist] nations” (Marx 1963: 52). This conception of humanity as not-yet-realized has at least two important implications for Marx's overall philosophy of the human being. First, like Hegel, Marx endorses a form of historicism according to which the actual lives of human beings are characterized differently at different times in history. Marx, along with Engels, has become extremely well known for his “dialectical materialism” or “historical materialism,” according to which *material* conditions of people at different stages determine the form of “human” life at that stage, shape the philosophies and “ideologies” that dominate that period, and also – inevitably – set the stage for a transition to a different stage in world history, one with different material conditions of labor and thereby different ideologies and forms of life. Second, unlike Hegel, Marx sees the historical process as far from finished, and this leads him to emphasize *action* over mere thought, revolution over mere philosophizing. The rest of this section briefly takes up these two themes.

Marx's dialectical materialism involves two central claims: first, that historical changes are results of an "inevitable" or "necessary" development (Marx 1963: 115, 120), and, second, that the fundamental basis for these changes are *material* conditions, where "material" here primarily refers to the control over the means for putting human work to use, whether these be physical factories or institutions or knowledge. Marx specifically distinguishes his philosophy of history from Hegel's in emphasizing not the unfolding of rational ideas into material relations but how material conditions generate the ideas of a particular epoch: "[c]onceiving, thinking, and the intellectual relationships of men appear here as the direct result of their material behavior . . . In contrast to German philosophy, which descends from heaven to earth, here one ascends from earth to heaven" (Marx 1970: 47, in Marx 1994: 111, see too Marx 1977: 102).

Marx's materialism is opposed to Hegel in an abstract way, but is also opposed to many popular conceptions of historical change. For Marx, ideas do not change the world, but vice versa. Only as the material conditions of production change can ideas change, since ideas are merely the reflection of those conditions: "Humans are the producers of their conceptions, ideas, etc., but these . . . humans . . . are conditioned by a definite development of their productive forces" (Marx 1970: 47, in Marx 1994:111). One does not change society by changing "hearts and minds;" rather, hearts and minds change as the conditions under which people live and work change. Similarly, historical change is not fundamentally a matter of *political* change. Because "legal relations [and] political forms . . . originate in the material conditions of life" (Marx 1994:210), any *fundamental* shift must always be in material conditions. Real political change must be the *consequence*, not the cause, of this material change.

Importantly, Marx's "materialism" is not the *physical* materialism of atoms in a void, but the *economic* materialism of concrete productive activities. Human beings must labor, first to survive, and then to express their species-being. And this labor is necessarily social in that we labor as members of community, not isolated individuals.

In the social production of their existence, men inevitably enter into definite relations, which are independent of their will, namely relations of production appropriate to a given stage in the development of their material forces of production. The totality of these structures of production constitutes the economic structure of society, the real foundation, on which arises a legal and political superstructure and to which correspond definite forms of social consciousness. The mode of production of material life conditions the

general process of social, political, and intellectual life. It is not the consciousness of men that determines their existence, but their social existence that determines their consciousness. (Marx 1994: 211)

Explaining the evolution of the material forces of production and thus the relations of production gets extremely detailed, and it is largely the effort to detail the important shift to a capitalist economy and the forces driving a shift *from* that economy that preoccupy Marx's intellectual life. But the general structure of the account involves looking at the ways in which human society at a given time is conditioned by the ways in which human work is capable of taking place, and then tracing how the development of processes of human labor create needs for new social, political, and economic structures. Thus, for example, Marx ascribes the difference between medieval-feudal and ancient economic systems to the "sparse population . . . scattered over a large area" with declining agricultural production and little trade (Marx 1994:109-10). Under such conditions, while the feudal system remained "an association directed against a subjected producing class[,] . . . the form of association and the relationship to the direct producers [was] different because of the different conditions of production" (Marx 1994:110). Marx's account of history, moreover, is neither an account of *individual* human beings acting under historically local conditions of production nor a description of humanity *in general*. As indicated in the example of the feudal system, Marx's focus is on associations of producers and those who exploit them, that is, on *classes*. Thus not only are ideas in a given period shaped by that period's material conditions, but, more specifically, "In every epoch, the . . . class that is the ruling *material* power of society is at the same time its ruling *intellectual* power" (MSW, 129). In general, "The history of all hitherto existing society is the history of class struggles" (Marx 1994: 158).

Ultimately, while Marx defends historical materialism in general, his focus is on applying it to his own capitalist society. He emphasizes how class struggles of previous ages – amongst "patricians, knights, plebeians, and slaves" or "feudal lords, vassals, guild-masters, journeymen, apprentices, and serfs" (Marx 1994:159) – are simplified into a single class antagonism between the bourgeoisie and the proletariat. The bourgeoisie own the (increasingly industrial-scale) means of production, while the proletariat work. We have already seen Marx's analysis of workers' alienation. In the context of dialectical materialism, Marx emphasizes how capitalism generates contradictions that will give rise to its own downfall. "The bourgeoisie not only has forged the weapons that bring death to itself; it has also called into

existence the men who are to wield those weapons — the modern working class — the proletarians” (Communist Manifesto, in Marx 1994:164).

Capitalism has at least three important effects that – Marx argues – will provoke a shift to a communist society. First, capitalism unites workers and gives them the tools of resistance. Mechanization makes all work essentially equal in worth and thereby makes workers more and more united. Improved means of communication (from telegraph to internet) make it easier for workers to work together as a class. And capitalists inadvertently arm workers with tools of organization through attempts to exploit them politically. Capitalists at first mobilize workers to use against competitors, as, for instance, when bosses of domestic industries incite workers to push for high import tariffs. But this political mobilization can (and, Marx argues, will) eventually backfire when workers recognize their solidarity with other workers around the world, rather than with their own industry and its bosses. Second, capitalism sows the seeds of its own destruction, not only through the alienation and *general* misery of the workers, but through sets of economic crises – what we might call the “business cycle” – that, Marx argues, become increasingly intense as mechanisms for mitigating their effects dry up. The result will be increasingly severe cycles of booms and busts in productive output and correspondingly severe cycles of unemployment, eventually leading to revolutionary demands for radical change in the economic structure. Finally, by increasing mechanical efficiency, capitalist society will change the structure of human need. Within capitalism, economic crises are caused not by *under*production but by *over*production; the failure to find sufficient markets for the excessive numbers of products leads to lay-offs and massive unemployment. But this opens an opportunity. Abolishing private property in a context of extreme scarcity could lead to severe deprivation of even basic needs. But when human beings can easily produce what they need through effective use of technology, most labor can be devoted towards work that is *not* tied to human needs. We can finally be truly *human* in our species-being. Thus capitalism creates a class that *could* bring about a shift to a new world order. It gives that class the *motive* to bring about that shift. And it makes such a shift *possible*: “The development of Modern Industry . . . cuts from under its feet [its] very foundation . . . What the bourgeoisie . . . produces, above all, are its own grave-diggers. Its fall and the victory of the proletariat are equally inevitable” (CM in Marx 1994:169).

How precisely this shift was to occur is a matter of some contention amongst Marxists. At the end of *The Communist Manifesto*, Marx appeals for a “communist revolution” crying out “Working men of

all countries, Unite!” In later works, Marx seems to allow that the “revolution” might be peaceful and political rather than violent and military. In any case, Marx often endorses the notion that the shift will not proceed directly from capitalism to pure, property-less communism but will proceed through various forms of socialism and pseudo-communism, such as the “universal private property” that seeks to equalize wages for all (and that Marx compares to “universal prostitution”) (Marx 1963:153; Marx 1994:69). And Marx recognizes the possibility of many false friends of communism, such as the “bourgeois socialism” that desires to “redress social grievances” in order to secure the continued existence of the bourgeois state” (Marx 1994:181) and the “utopian socialism” that is little more than an idle side-project of the wealthy and a distraction from real social change (Marx 1994:182-4). In terms of Marx’s conception of human nature, however, the important point is simply that for Marx, all history prior to the achievement of a communist condition within which alienated labor will be overcome is merely the “prehistory of human society” (Marx 1994:212). Human beings *will be* the result of millennia of labor that develops towards the overcoming of class antagonism and the fulfillment of truly human potential.

Marx’s account of ideal human nature, alienation, and dialectical materialism would not, in themselves, have made him one of the defining figures of the 19th and 20th centuries. But Marx’s philosophy was not the complacent thought-experiment of an armchair philosopher. Instead, Marx put his ideas to use to mobilize, organize, and inspire the concrete activities of communists seeking to change that world. For Marx, “the chief defect of all hitherto existing materialism,” and of all previous philosophy, is that it “does not grasp the significance of ‘revolutionary’, of ‘practical-critical’, activity”: “The philosophers have only interpreted the world, in various ways; the point is to change it” (*Theses on Feuerbach* ##1, 11; in Marx 1994:99, 101). In describing the dialectically necessary progress towards communism in the Communist Manifesto, Marx applies this practical-critical emphasis to a thinly veiled autobiography of his role in historical progress:

[I]n times when the class struggle nears the decisive hour, the progress of dissolution . . . assumes such a violent, glaring character, that a small section of the ruling class cuts itself adrift and joins the revolutionary class, the class that holds the future in its hands. Just as . . . at an earlier period, a section of the nobility went over to the bourgeoisie, so now a portion of the bourgeoisie goes over to the

proletariat, and in particular, a portion of the bourgeois ideologists, who have raised themselves to the level of comprehending theoretically the historical movement as a whole. (CM in Marx 1994:167)

Through his philosophical descriptions of alienation and even more through his political-economic analyses of the failures of capitalism to meet basic human needs, Marx provides the theoretical framework for uniting the proletariat in its revolutionary struggle.

This practical focus manifested itself in Marx's writing style, which tended to be polemical rather than philosophical, and in his active involvement in worker's movements around Europe. The ultimate result was that Marx's ideas – albeit in warped and incomplete forms – shaped the ideology of two of the three most powerful nations in the world during the 20th century, and even today over 1 billion people live under (nominally) “Communist” rule. Today, after the fall of the Soviet Union and the increasingly capitalist economic structure in China, Marxism in its Leninist-Stalinist-Maoist form is increasingly discredited. But the end of these *supposed* communist systems opens the possibility for a reappraisal of the contributions that *Marx* – as opposed to *Marxism*¹⁶² – can and has made to our conception of human beings. For example, while Marx's strictly *political* influence may be waning, his *intellectual* influence is still significant in disciplines such as history and sociology, where not only Marx's general emphasis on material conditions of human life but even many of his specific analyses continue to exercise significant influence. More generally, although Marx is widely disclaimed by professional economists in favor of theorists such as Smith or Ricardo, it was Marx – far more than Smith or Ricardo – who argued that economics is the *fundamental* science of human nature, and the increasing dominance of economic ways of conceiving of human beings can trace its origin, at least in significant part, to Marx. Finally, even if Marx's description of alienation and his utopian vision for a communist world order are overstated, his conception of human beings as needing to find a fulfillment in their lives that capitalist economic structures preclude is a conception that needs to be taken seriously in a world in which ever more wealth seems poorly correlated with richer and fuller human lives.

II. Darwin and the Rise of Biology

¹⁶² See the aptly titled *Marx after Marxism* (Rockmore 2002).

When Kant was working through his philosophy of biology in the *Critique of Judgment*, Newton's *Principia* had established a pattern for scientific physics, but the closest parallel in biology was Buffon's *Natural History*, which consisted primarily in the classification of various different species based on their anatomical characteristics. Biology as a discipline that could *explain* the nature of living things was still in its infancy. The new physics had, for a short time, given hope to purely mechanistic accounts of biology, according to which one could see living things as complex machines. But in Kant's day, these mechanistic approaches to biology had given way to either various forms of "preformationism" or vitalist, quasi-mechanist approaches to the origin of life. Preformationism posited that each species was separately created and that individual members of those species were present, in some sense, in their ancestors. In its crudest version, this implies that every human being was present, in miniature, in the eggs of Eve. Vitalist explanations were closer to mechanist ones, positing that complex forms of life emerge from simpler material interactions, but vitalists posited that matter itself was best understood on a biological rather than strictly mechanistic model. (Schelling's attempt to infuse nature with something analogous to freedom is an example of how this vitalism developed in the early 19th century.) The challenge for all of these forms of biological explanation was how to make sense of the apparent *purposiveness* of living things; our hearts seem best explained by the fact that we need blood to be pumped through our bodies, for example. As we saw in chapter one, Kant's own philosophy of biology affirms this purposive sort of explanation as basic to biology while according it a merely "regulative" status. But while Kant holds open the possibility in principle of synthesizing these purposive explanations of organisms with a mechanistic account of their origins, he doubts the emergence of a "Newton of a blade of grass."

Arguably, Charles Darwin (1804-1882) is just such a biological Newton. He provided the basis for a new sort of mechanism in biology that showed how purposive structures of organisms can emerge from natural and not intrinsically purposive causal processes. Over time and under various environmental influences, random variations in organisms give rise to increasingly refined and even purposive structures through a process of "natural selection." This new approach put biology on scientific footing by establishing an intuitively plausible and rigorous methodology for not merely classifying organisms but also explaining their origins and variations. But Darwin's new theory of evolution by natural selection all-too-clearly applied to human beings as well. As natural organisms, our own specific characteristics must have emerged

from a natural process of evolution by natural selection. For many, this offered new hope of finally and decisively answering the question “What is the human being?” For others, it offered reason for despair, seemingly making humans no more than animals and precluding the possibility of any of those characteristics – such as reason, freedom, and the capacity for love – that make human beings so unique.

Darwin’s theory of evolution by natural selection, as described in his most famous work – *The Origin of Species* – is fairly simple. Biological reproduction has two important characteristics: first, descendants tend to share most of the traits of their ancestors; and second, organisms tend to reproduce in sufficiently large numbers that there is a “struggle for existence,” by virtue of which many descendants die early or are unable to reproduce. Darwin argues,

Owing to the struggle for life, any variation, however slight and from whatever cause proceeding, if it be in any degree profitable to an individual of any species, . . . will tend to the preservation of that individual, and will generally be inherited by its offspring. The offspring, also, will thus have a better chance of surviving . . . I have called this principle, by which each slight variation, if useful, is preserved, by the term of Natural Selection. (Darwin 1859/1985: 115)

Over long spans of time, preserved variations tend to become more prevalent within a population, and the species “evolves.” Moreover, given that there is often more than one way members of a species can thrive, “the greatest amount of life can be supported by the greatest diversification of structure” (Darwin 1859/1985: 157). If different members of a population have different variations that are valuable for different purposes, the species diverges, eventually becoming two or more distinct species.

The power of this approach is evident in the title of Darwin’s work: *On the Origin of Species*. Over the course of the 18th and early 19th centuries, the idea that one could provide an account of the origin of biological entities in terms of mechanical forces had increasingly lost ground against biological theories that took for granted an original creation of the full range of biological species, each with their distinctive traits firmly established. In part, this view in biology was motivated by religious commitments, but in part, it was also due to the apparent failure to find any clear mechanism by means of which complex biological interdependence could be explained. As Kant laid out the case in his *Critique of Judgment*, organisms seem to be *teleologically ordered*, both in terms of physiology (the heart seems to exist for the *purpose* of

pumping blood for a body on which it is itself dependent) and in terms of ecology (lions need gazelles to eat and gazelles need lions to control their populations). But Darwin's straightforward account made perfect sense of how species could evolve, change, and diverge through selection processes that are not, in themselves, purposive. In that sense, it revolutionized biology, making both special creations and inherent teleology obsolete.

In his *Origin of Species*, Darwin barely mentions *human* evolution, saying only

In the distant future I see open fields for far more important researches. Psychology will be based on a new foundation, that of the necessary acquirement of each mental power and capacity by gradation. Light will be thrown on the origin of man and his history. (Darwin 1859/1982: 458)

Still, the implications of *The Origin of Species* for the case of human beings were immediately obvious to both Darwin and his contemporaries. In notebooks containing writings from a period during which he was first developing his theory of evolution, Darwin points out "Man . . . is not a deity," challenges those who "dare boast of [humans'] preeminence," considers "What circumstances may have been necessary to have made man" and even compares human beings to orangutans (Darwin's Notebook C, §77-79, written in 1838, Darwin 1987:263-4). And in 1863 – just 5 years after the first edition of *Origin* – T.H. Huxley began his *Evidence as to Man's Place in Nature* with a diagram of the human skeleton juxtaposed to the apes to which Huxley argued he was related.

Darwin entered the fray most explicitly with the publication of his *The Descent of Man* in 1871. In this work, Darwin makes use of the increasingly wide acceptance of the notion that there are homologous structures in man and the lower animals: "man is constructed on the same general type or model with other mammals" (Darwin 1871/1902: 10). Like dogs and even pigeons, human beings have a heart, two lungs, a skeletal structure including vertebrae, a skull, ribs, knees, and even two distinct bones (radius and ulna) in the fore-arm/leg/wing. Darwin uses this physical similarity between man and other animals for two important and related purposes. First, it provides an important piece of "evidence of the descent of man from some lower form" (9). Second, and arguably even more importantly, Darwin uses the model of homologous *physical* structures as an analogy for homologous *mental* structures. Darwin argues both that "there is no *fundamental* difference between man and the higher mammals in their mental faculties" (Darwin

1871/1902: 35, emphasis added) and that – therefore – the human mind/soul is descended from some lower form just as much as the human body. This shift from physical to mental homology is radical: recall that Descartes, long before Darwin and without raising any significant threat for religion, had argued that the human *body* could be explained as the natural result of evolution by means of natural, mechanical interactions of matter in motion. But Descartes reserved a special place for the human *soul*, which, he claimed, was uniquely able to explain humans’ rational superiority over animals. Likewise today, many who are quite comfortable with the idea of humans’ physical nature as naturally evolved nonetheless argue that there is something special about humans requiring a separate creation.

Darwin would have no part in this idolizing of human beings. He was, of course, insistent that “the difference in mental power between the highest ape and the lowest savage [is] immense” (34). But he insisted that animals share many supposedly “human” traits, including reason, abstraction, and even ennui and a capacity for deceit (Darwin 1871/1902: 109, 117f., 102, 99). And, more importantly, he offered an evolutionary account of the origin of humans’ most distinctive traits, including language, higher cognitive faculties, morals, and even religious belief. Among key aspects of his account are his evolutionary treatments of language, religious belief, and a sense of beauty. But the most important aspect of his evolutionary account of human beings is Darwin’s discussion of the evolution of humans’ “moral sense” (Darwin 1871/1902: 134ff.). Darwin starts with a sympathy-based (broadly Humean or Smithian) moral theory:

any animal whatever, endowed with well-marked social instincts . . . would inevitably acquire a moral sense or conscience, as soon as its intellectual powers had become as well, or nearly as well developed, as in man. For, firstly, the social instincts lead an animal to take pleasure in the society of its fellows, to feel a certain amount of sympathy with them, and to perform various services for them. . . . Secondly, as soon as the mental faculties had become highly developed, images of all past actions and motives would be incessantly passing through the brain of each individual: and that feeling of dissatisfaction, or even misery, which invariably results, as we shall hereafter see, from any unsatisfied instinct, would arise, as often as it was perceived that the enduring and always present social instinct had yielded to some other instinct, at the time stronger, but neither enduring in its nature, nor leaving behind it a very vivid impression Thirdly, after the power of language had been acquired, and the wishes of the community could be expressed, the

common opinion how each member ought to act for the public good, would naturally become in a paramount degree the guide to action. But it should be borne in mind that however great weight we may attribute to public opinion, our regard for the approbation and disapprobation of our fellows depends on sympathy, which, as we shall see, forms an essential part of the social instinct, and is indeed its foundation-stone. (Darwin 1871/1902:135-6)

Darwin then makes use of his evolutionary account of the origin of language and reasoning, to fill in the cognitive part of this process and provides an evolutionary argument for the sympathy that is its “foundation-stone.” As in the rest of his account of human beings, Darwin emphasizes the similarities between human beings and other animals, noting that humans are only one of a wide range of social animals, and appealing to dogs in particular as examples of animals with social instincts including love, sympathy, and self-command, even to the point of “something very like a conscience” (Darwin 1871/1902: 142). Having shattered the uniqueness of humans’ social affections, Darwin proceeds to offer a general account of the advantages of social instincts for the groups whose members have them, and thus of the likelihood that those instincts will be passed on to offspring. Darwin even specifically quotes, at the beginning of his discussion of human morals, *Kant’s* description of our moral predisposition: “Duty! Wondrous thought . . . before whom all appetites are dumb” (Darwin 1871/1902: 134-5). Rather than taking this as an inexplicable human given, though, Darwin explains its origin through his principle of evolution through natural selection.

The implications of Darwinism for philosophy in general have been profound. Within Descartes’ dualistic conception of the human being as having a mechanistic body and a free immaterial mind-soul, philosophical discussions of the nature of human beings could focus on the human mind and thereby insulate themselves from the rise of scientific accounts of the body. Alternatively, Romantic biology and its Schellingian-Hegelian variations made even physical biology fundamentally vitalist in a way that required philosophical reflection on the purposes of nature. But Darwin provided a methodology for an empirical biology that marginalized many sorts of reflections on human nature that had dominated philosophy. Moreover, the scope of Darwin’s theory seemed to include every aspect of human nature, leaving nothing for philosophers of human nature to do except engage in Darwinian biological research.

For Kant, Darwinism might seem particularly problematic. At the most mundane and direct level, Darwin seemingly provided a refutation of Kant's philosophy of biology. Kant presumed that teleology was intrinsic to the study of biological entities as such, but Darwin showed how *apparent* teleology is merely the effect of a process of evolution by natural selection that is not teleological at all. This rejection of teleology at the foundation of biology could call into question the unification of theoretical and practical reason that Kant sought to effect in his *Critique of Judgment*. But Darwinism might also seem to pose even more serious problems, since it suggests that the transcendental anthropology at the core of Kant's answer to the question "What is the human being?" is fundamentally secondary to a revamped empirical anthropology. And even within Kant's empirical anthropology, Darwin suggests that the most important part of any such anthropology is precisely the part that Kant ignored: the description of how humans' "predispositions" (in Kant's lingo) evolved from more primitive structures. Only such study can provide real explanations for human nature, and given these explanations, one can better understand the nature and role of the faculties that Kant erroneously sought to explore through transcendental philosophy. Moreover, the recognition of the way in which such structures evolved will require giving up the idea that any human mental powers are *fundamentally* fixed.

Of course, Darwin's theory lacked sufficient evidence at the time, and he left many issues unresolved, such as the sources of variations amongst individuals – which, Darwin said, "seem to arise from . . . unknown causes" (Darwin 1871/1902: 97) – and the specific biological mechanisms by which advantageous traits were passed on to offspring. But gaps in evidence were relatively insignificant in comparison with the added explanatory power that it provided. The unresolved issues were potentially more devastating in the early 20th century when the results of Gregor Mendel's work on genetics showed that traits were not heritable in the way that Darwinian evolution seemed to require.¹⁶³ But since 1953, when Watson and Crick discovered the molecular basis for genetics in the DNA molecule, Darwinians have embarked upon a "modern synthesis" of Darwinian selection and molecular genetics, one that aims to provide an integrated and complete evolutionary account of human beings.

¹⁶³ For discussion of these issues, their resolution, and some remaining issues for Darwinism in the 21st century, see Depew and Weber 1996 and Kitcher 1984.

III. Nietzsche, Art, and Literature

While Darwin and Marx helped shift the question “What is the human being?” outside of the field of philosophy, Friedrich Nietzsche (1844-1900), though trained as a philologist rather than a philosopher, has become the most well-known and widely read 19th-century *philosopher* of human nature. Unlike Darwin and Marx, however, Nietzsche refuses to advance a particular answer to the question. His philosophy of human nature can be seen as a continuation of the Romantic appropriation of and response to Kant that we examined in the last chapter. Like the Romantics, Nietzsche is anti-systematic and anti-metaphysical, even employing the same aphoristic form cultivated by the Romantics. Thus at the end of one of his most important articulations of his philosophy, Nietzsche laments,

Alas, what are you after all, my written and painted thoughts! It was not long ago that you were still so colorful, young, and malicious, full of thorns and secret spices . . . and now? You have already taken off your novelty, and some of you are ready, I fear, to become truths: they already look so immortal, so pathetically decent, so dull! (BGE § 296, Nietzsche 1966: 236¹⁶⁴)

While Nietzsche articulates provocative and important claims about human nature, his writings are also permeated with a Romantic resistance to theory that precludes his settling on any particular “theory” of human “nature.” Nonetheless, important claims about human beings can be gleaned from his writings, claims that he repeats, develops, and reiterates, even if he would resist them being classified as “Nietzsche’s answer” to Kant’s question.

Like the German Romantics, Nietzsche stresses creativity and individuality over abstract rationality and universal duty. He pulls these Romantic themes together with a sort of historicism – akin to Herder’s – that emphasizes the contingency and a-rationality of the particular systems of thought and morals that exist in the present. And Nietzsche situates all of this in the context of an optimistic Schopenhauerian conception of the will to power as an underlying creative force in the universe that seeks ever higher forms of expression through a self-overcoming that always involves suffering and, at its best, involves a *creative* suffering, “like pregnancy” (GM §2.19, Nietzsche 1967: 88). In

¹⁶⁴ Throughout, references to Nietzsche provide both section numbers from particular works and the page numbers in recent translations. Abbreviations for Nietzsche’s works are BGE for *Beyond Good and Evil*, Z for *Thus Spake Zarathustra*, GM for *On the Genealogy of Morals*, GS for *The Gay Science*, and EH for *Ecce Homo*.

relation to Kant, it will be helpful to emphasize four of Nietzsche's primary contributions to thinking about human beings: (1) his genealogical methodology, whereby human cognition and morals are seen as contingent historical-cultural perspectives; (2) his opposition to certain metaphysical conceptions of the self, especially those involving a coherent and unified "I" or a Kantian conception of human "freedom"; (3) his perspectivism, which draws from but radicalizes Kant's Copernican turn; and (4) his conception of the "overman," the idea that our current configuration of moral and philosophical perspectives can and should be overcome and that a new, higher type of human being can emerge.

Nietzsche's impact has been most influential in thinking about morals, and the title of his *Genealogy of Morals* – much like Darwin's *Origin of Species* – highlights the profound shift he aims to inaugurate. Unlike the "stiff seriousness that inspires laughter" of "all our philosophers" who "wanted to provide a *rational foundation* for morality," Nietzsche will

own up in all strictness to what is still necessary here for a long time to come . . . : to collect material, to conceptualize and arrange a vast realm of subtle feelings of value and differences of value which are alive, grow, beget, and perish—and perhaps attempts to present vividly some of the more frequent and recurring forms of such living crystallizations—all to prepare a *typology* of morals. (BGE § 186, Nietzsche 1966: 97)

Just as natural scientists required Darwin to shake them from confidence in the fixity of species, so moral philosophers can thank Nietzsche for shattering their assumption of a given, fixed morality for which they could provide the "conditions of possibility." As Nietzsche puts it, "my curiosity as well as my suspicions were bound to halt quite soon at the question of where our good and evil really *originated* . . . [U]nder what conditions did man devise these value judgments good and evil" (GM §Preface.3, Nietzsche 1967: 17).

It is not necessary to get into the details of Nietzsche's account of the origin of 19th century European values here. Throughout, he appeals to social and natural forces in explaining shifts in human values, pointing out, for example, how "the change which occurred when [man] found himself finally enclosed within the walls of society and of peace" gave "old instincts" a new – inward – direction (GM §2.16, Nietzsche 1967: 84, cf. BGE § 201). The most important shift, for Nietzsche, involved turning away from a "noble morality" that was fundamentally self-affirming, active, and strong, the morality of the ancient and heroic

Greeks, the morality of Achilles or – even earlier – Gilgamesh. For Nietzsche, this morality was overturned in a “slave revolt” in morals, a context where the weak and oppressed turned against their oppressors, not through active revolt but through a subtle “revaluation of values” that rejected as “vices” the strengths of the nobles and affirmed as “virtues” the characteristics of the weak. As Nietzsche so eloquently puts it, “Weakness is being lied into something *meritorious* . . . and impotence which does not requite into ‘goodness of heart’; anxious lowliness into ‘humility’; subjection to those one hates into ‘obedience’” (GM §1.14, Nietzsche 1967: 47). Kant’s emphasis on autonomy is recast as an “instinct of obedience” to “formal conscience” (BGE §199, Nietzsche 1966: 110), a slavish and herdlike “morality as timidity” (BGE §197, Nietzsche 1966: 109). Against Kant’s “bad taste of wanting to agree with many”, Nietzsche proclaims that “My judgment is *my* judgment; no one else is entitled to it.” “How,” he asks, “should there be a ‘common good’ . . . [W]hatever is common always has little value” (BGE § 43, Nietzsche 1966: 53). But as important as these descriptions of Kant’s morality as common, timid, and slavish are, Nietzsche’s shift from *justifying* a timeless morality transcendentally to describing the emergence, changes, and possibilities of contingent moralities is – as we will see in more detail in chapter nine – Nietzsche’s most lasting and influential “criticism” of Kant.

For Nietzsche, the shift from justification to genealogy, and the exposing of supposed philosophical truths as merely natural phenomena, is not limited to morality:

Perhaps the time is at hand when it will be comprehended again and again *how little* used to be sufficient to furnish the cornerstone for such sublime and unconditional philosophers’ edifices as the dogmatists have built so far: any old popular superstition from time immemorial (like the soul superstition, which, in the form of the subject and ego superstition, has not even yet ceased to do mischief); some play on words, perhaps, a seduction by grammar, or an audacious generalization of some very narrow, very personal, very human, all too human, facts. (BGE, Preface, Nietzsche 1966:1)

The whole edifice of philosophical truths is built, for Nietzsche, on contingent and historically emergent prejudices. In criticizing these “philosophical truths,” Nietzsche offers general criticisms of metaphysical systems, such as that “every great philosophy so far has been . . . the personal confession of its author and a kind of unconscious and involuntary memoir . . . [in which] the moral (or immoral) intentions . . . constitute [its] real germ of life” (BGE §6, Nietzsche 1966: 13). And

Nietzsche also takes aim at concepts particularly important for Kant's philosophy. About the fundamental question of Kant's *Critique of Pure Reason*, Nietzsche says,

[I]t is high time to replace the Kantian question, "How are synthetic judgments *a priori* possible?" with another question, "Why is belief in such judgments *necessary*?" – and to comprehend that such judgments must be *believed* to be true, for the sake of the preservation of creatures like ourselves, though they might, of course, be *false* judgments for all that! (BGE § 11, Nietzsche 1966: 19).

Rather than a transcendental justification of the legitimacy of such judgments, Nietzsche challenges his readers to think of the interests that are served by them, and thus undermines the whole foundation of Kant's *Critique*. With respect to freedom, Nietzsche's opposition to Kant is more specific. First, he argues that while "[p]hilosophers are accustomed to speak of freedom as if it were the best-known thing in the world," in fact, this approach only "adopt[s] a *popular prejudice* and exaggerate[s] it" (BGE § 19, Nietzsche 1966: 25). The whole concept of "freedom" is "a sort of rape and perversion of logic" and "nonsense" (BGE §21, Nietzsche 1966: 28). Against this empty, formal, nonsensical notion of freedom, Nietzsche offers an alternative, wherein the "will" is a "complex of sensation, thinking, and above all *affect*," such that "freedom" is always the freedom *of* one aspect of human nature to dominate others (BGE § 19, Nietzsche 1966: 25). Kant's "transcendental freedom" is the domination of a particular instinct over others, rather than – as Kant supposed – a freedom from domination by instinct altogether. Once willing is properly understood as "something complicated," the sort of metaphysical freedom appealed to by Kant is exposed for what it is, an empty verbal embrace of an incoherent prejudice.

For Nietzsche, this general unmasking of philosophical pretensions to "absolute truth" or "universality" was only part of a more general effort to radicalize and relativize Kant's Copernican turn in the service of a perspectivism that would in turn pave the way for a more creative approach to human existence. Nietzsche asks "under what conditions did man devise value judgments good and evil?" only to go on to ask "*and what value to they themselves possess? Have they hitherto hindered or furthered human prosperity?*" (GM §Preface.3, Nietzsche 1967: 17). What Kant saw as necessary conditions of the possibility of any human experience become, in Nietzsche's hands, particular prejudices of

particular ages, embodied in language and shared prejudices. In the context of distinguishing between mere “philosophical laborers” and true “philosophers,” Nietzsche explains the real value (for him) of mere laborers like Kant.

It may be necessary for the education of a genuine philosopher that he himself has once stood on all the steps on which . . . the scientific laborers of philosophy remain standing . . . in order to pass through the whole range of human values and value feelings and to be *able* to see with many different eyes and consciences . . . Those philosophical laborers after the noble model of Kant and Hegel have to determine and press into formulas, whether in the realm of *logic* or *political* (moral) thought or *art*, some great data of valuations—that is, former *positings* of values, creations of value which have become dominant and are for a time called “truths.” It is for these investigators to make everything that has . . . been esteemed so far easy to . . . think over, intelligible and manageable . . . *Genuine philosophers, however, are commanders and legislators: they say, “thus it shall be!”* . . . With a creative hand they reach for the future, and all that is and has been becomes a means for them, an instrument, a hammer. (BGE § 211, Nietzsche 1966: 136)

Kant’s transcendental analyses provide an invaluable deep and detailed account of the conditions of possibility of particular ways of thinking, feeling, and acting. But Nietzsche’s genealogies do more. They pave the way for a truly creative *revaluing of values*, one within which particularity, creativity, strength, and life would be affirmed.

Becoming conscious of the historical contingency of one’s values and prejudices opens a new sphere of freedom, a recognition that values and prejudices are precisely *not* an a priori that constrains us but a set of tools to be utilized *as we see fit*. And insofar as we live – as Kant argues – in a world that is in part the product of our presuppositions and values, our power *over* these presuppositions is a power to create *new worlds*: “it is enough to create new names and valuations to create new ‘things’” (GS §58, Nietzsche 1974: 122). There is, of course, something frightening about this freedom.

We have forsaken the land and gone to sea! We have destroyed the bridge behind us – more so, we have demolished the land behind us! Now, little ship, look out! Beside you is the ocean . . . [T]here will be hours when you realize that it is infinite and there is nothing more awesome than infinity. Oh, the poor bird that has felt free and now strikes against the walls of this cage! Woe, when homesickness for

the land overcomes you . . . and there is no more 'land'! (GS § 124, Nietzsche 1974: 180-1)

Having seen the contingency of all values, it is no longer possible to go back to the naïve “a priori” to which we had only to submit. But homesick nostalgia for unreflective naïveté is not the only possible reaction to the death of our old prejudices and values, our old “God.”

[A]t hearing the news that ‘the old god is dead’, we philosophers and ‘free spirits’ feel illuminated by a new dawn; our heart overflows with gratitude, amazement, forebodings, expectation—finally the horizon seems clear again, even if not bright; finally our ships may set out again, set out to face any danger; . . . the sea, *our* sea, lies open again; maybe there has never been such an ‘open sea’. (GS § 343, Nietzsche 1974:280)

For Nietzsche, the bases for our old values and morals are gone. We have moved – or at least, should move – beyond good and evil. Some will respond to this loss with despair. Most will respond with a self-deception that refuses to admit the loss. Pretending that the old values still live, they will throw themselves into business and “commonsense” to avoid the reality that they are living on an open ocean. But this openness gives rise to a new sort of ideal, an ideal of the true “philosopher” and “free spirit,” the one who can respond to the loss of naïveté with honesty, courage, and the strong creativity needed to form *one’s own* values.

This emphasis on creativity, on the possibility of new possibilities for human beings, makes Nietzsche a philosopher *of the future*¹⁶⁵. For Nietzsche, “the human being is something that shall be overcome,” a mere “rope” between “beast” and what Nietzsche calls *der Übermensch*, literally that which is over, or beyond, the human being (Z, Nietzsche 1978: 12, 14). Given Nietzsche’s condemnation of universality, there is not – and cannot be – a formula for what an *Übermensch* is. But Nietzsche does lay out a few general visions of a future, better, humanity. An *Übermensch*, unsurprisingly, will be creative and self-confident, not seeking to accommodate his views to those of the masses but willing to strike out on his own. This self-confidence goes further in Nietzsche’s doctrine of the “eternal recurrence.” In describing *Thus Spake Zarathustra*, Nietzsche claims that “the idea of the eternal recurrence” is “the fundamental conception of this work” because this notion is “the highest formula of affirmation that is at all attainable” (EH, Nietzsche 1967: 295). The thought that the world will repeat itself infinitely, that

¹⁶⁵ See the subtitle to *Beyond Good and Evil: Prelude to a Philosophy of the Future*.

“whatever was and is” will be “repeated into all eternity” (BGE § 56, Nietzsche 1967: 68) is a thought that at first terrifies Zarathustra (the protagonist of *Thus Spake Zarathustra*), but which ends up being embraced in those moments when Zarathustra is most akin to the *Übermensch*. The *Übermensch* will be the “opposite” of all of those life-denying and pessimistic moralists; he is an “ideal of the most high-spirited, alive, and world-affirming human being . . . shouting insatiably *da capo* [repeat] – not only to himself but to the whole play and spectacle [and] who makes [this whole play] necessary because again and again he needs himself—and makes himself necessary” (BGE § 56, Nietzsche 1978: 68). The ideal person is so self-affirming that he is willing to affirm all of the pettiness, misery, and evil in the world, and to affirm that again and again, on the grounds that all of this was and is worth it because it went into making a world that included *himself*.

But this self-affirmation also implies that the *Übermensch* live her life in such a way that it *can* justify the world. “*Übermensch*” takes the place of the “God” who is now dead and the moralities of mediocrity that lived off God for centuries. While “once one said ‘God’ when one looked upon distant seas,” Nietzsche now invites us to say “*Übermensch*” (Z, Nietzsche 1978:85). And for Nietzsche, there is a crucial difference between these ideals: “God is a conjecture, but I do not wish your conjecturing to reach beyond your creating will. Could you *create* a God?— Then, I pray you, be silent about all gods! But you could well create the *Übermensch*” (Z, Nietzsche 1978:85). Nietzsche is ambivalent about the *full* extent to which one can bring about the *Übermensch* – the preceding quotation continues “Not perhaps you yourselves . . . but into . . . forefathers of the *Übermensch* you could transform yourselves” – but the *Übermensch* is, at least, a goal that is attainable in principle, something towards which we can orient our active, creative powers rather than someone to whom to submit. Thus Nietzsche (in the guise of Zarathustra) asks, “What have *you* done to surpass humankind?” (Z, Nietzsche 1978:12, my emphasis).

Unlike Darwin, Marx, and (as we will see) Freud whose legacy is clearly identifiable in contemporary biology, psychology, and social sciences, Nietzsche did not leave behind a discipline distinct from philosophy. But Nietzsche’s legacy lingers today not only in the discipline of philosophy itself, where he continues to be one of the most widely read philosophers, but – perhaps more importantly – in the ever increasing emphasis on art, literature, and now film as sources for answering the question, “What is the human being?” In fact, just as some may be likely to turn to biology or psychology as *sciences* that can answer that question, those with a sense that there is more to being human than what science can capture are at least as likely – if not more likely – to

turn to literature and art than to philosophy. And this emphasis on literature as a source for thinking about humanity reflects deeply Nietzschean impulses about the nature of that question. For one thing, literature refuses to simplify human behavior into simply formulae or universal rules; “common” literature is bad literature, and that which is truly great is something particular, individual, *extraordinary*. Moreover, good literature exposes its readers to a range of human possibilities, opening new vistas and perspectives rather than simplifying all perspectives into a taxonomy. And good literature (or art) is precisely the literature worth reading and rereading, literature depicting the lives that meet Nietzsche’s vision of the *Übermensch* who can reflect with pleasure on the eternal return of all things because these are vindicated in her own interesting, original, dynamic, and creative life.

IV. Freud and the Rise of Psychology

While Marx helped inaugurate the rise of history and social sciences in general and Darwin brought new importance to biological studies of human beings, Freud can be seen as one of the key figures¹⁶⁶

¹⁶⁶ Arguably, Wilhelm Wundt, though much less well-known than Freud, was at least as important in the emergence of psychology as a viable science. Wundt was the first modern psychologist to set up a lab, and he developed a systematic method for investigating human mental states that helped get “psychology” as a discipline off the ground.

Wilhelm Wundt was born in 1832 in Mannheim and studied medicine at Tübingen and Heidelberg, where he remained until 1874, spending most of this period working alongside Hermann von Helmholtz, whose approach to human psychology was heavily influenced by both Kant and German empiricists. In 1874, Wundt took a chair in “inductive philosophy” at the University of Zürich, and the following year was appointed to a chair in philosophy at Leipzig. There Wundt set up a laboratory in experimental psychology and set to work developing an experimental approach to the study of the human mind. Wundt published extensively on a wide range of topics in psychology and philosophy, he had over 160 doctoral students work in his psychology laboratory, and he is widely credited with being the founder of modern experimental psychology.

Wundt’s most important impact on the history of psychology was sociological rather than philosophical. His laboratory-based psychological methodology and his large number of students were significant in helping to establish experimental psychology as a viable new discipline at the end of the 19th century. Theoretically, Wundt helped establish psychology as a rigorous experimental science of purely mental phenomena. Like Kant, Wundt focused on “special laws for our psychical life [that] . . . differ from the universal physical ones” (Wundt 1912: 154), but unlike Kant, he insisted that psychology is not limited to “the description of facts” but can attain the same level of “universally valid rules” as the natural sciences (Wundt 1912: 156).¹⁶⁶ Wundt articulated a variety of psychical laws, such as “the principle of intensifying contrasts,” shown in the fact that a line surrounded by smaller lines will look larger than the same line seen alone. And he developed a general theory of the basic “simple elements” of consciousness and their combinations as a way of showing how relations amongst complex psychic phenomena can be explained in terms of laws governing their simple elements. Wundt also uses his account of psychologically simple

behind the rising importance of *psychology* as a way of answering the question, “What is the human being?” In some respects, Freud can be seen as making Nietzsche’s emphasis on unconscious drives and genealogies of morals into a rigorous psychological science distinct from biology. At the beginning of the 20th century, Freud said:

You have been trained to find an anatomical basis for the functions of the organism and their disorders, to explain them chemically and physically and to view them biologically. But no portion of your interest has been directed to psychical life, in which, after all, the achievement of this marvelously complex organism reaches its peak. For that reason psychological modes of thought have remained foreign . . . This is the gap which psycho-analysis seeks to fill. It tries to give psychiatry its missing psychological foundation . . . With this aim in view, psycho-analysis must keep itself free from any hypothesis that is alien to it, whether an anatomical, chemical, or physiological kind, and must operate entirely with purely psychological . . . ideas. (Freud 1920/1963: 23-4)

Freud here articulates a general mood at the beginning of the 20th century, one that helped launch psychology, sociology, history, and even “anthropology” (which in today’s English means something quite different from what it meant in Kant’s German) as new realms of “human sciences” not reducible to mere biology. This quotation makes it seem as though there is a sharper contrast between psychology and biology than exists today, and sharper than Freud intends. But while Freud admits that psychological structures will ultimately be found to be based on biological structures (see, e.g. Freud 1920/1963:315-6 and Freud 1914), his more important point is that there is room for sciences that study human beings empirically in ways that are not reducible to physical or biological investigations. And these “human sciences,” among which

elements to highlight the distinction between psychological and physical laws, noting that while “there is no psychical process . . . which does not run in parallel with a physical process . . . , a simple [psychological] process . . . does not correspond to even a relatively simple . . . physical one” (Wundt 1912: 186).¹⁶⁶ Wundt also continually refined his methodology, shifting away from brute introspection and towards a more rigorously experimental method that isolated simple mental processes as a basis for further carefully constructed experiments to get at more complex phenomena. (In his *Introduction to Psychology* (Wundt 1912), for example, the auditory cognition of patterns of beats on a metronome gets dozens of pages of discussion.) Ultimately, Wundt provided a theoretically justified methodology for experimental psychology, a set of theories and concepts further psychologists could develop and correct, and a cohort of well-trained experimental psychologists who would go on to help found an autonomous discipline of empirical psychology that could compete with and/or supplement philosophy and biology in providing answers to the question, “What is the human being?”

psychology still dominates today, have transformed our conceptions of human beings.

Sigmund Freud (1856-1939), the inventor of psycho-analysis, is perhaps best known for a whole set of popular theories and expressions, such as “Freudian slips,” penis-envy, the Oedipal complex, the thesis that all neuroses can be traced to childhood sexuality, and, more generally, a basic obsession with sex in mental life. But Freud played an important part in the emergence of modern empirical psychology. He insisted that psychology should be an empirical science, emphasizing that “the psycho-analytic view is . . . empirical – either a direct expression of observations or the outcome of a process of working them over” (*Introductory Lectures*, p. 302) and comparing psycho-analysis with astronomy, as two subjects in which “experimentation is particularly difficult” but which can nonetheless make inferences based on observations (see, e.g., *New Introductory Lectures*, p. 27). Moreover, Freud emphasized the distinction between psychology and physiology and thereby helped provide a realm of its own to scientific psychology. Importantly, Freud also emphasized the clinical and therapeutic importance of non-physiological psychology and helped establish a model for productive interaction between clinical and theoretical work in psychology.

But Freud’s most important contribution, not only to modern empirical psychology but also to modern conceptions of the human being more generally, was his emphasis on *unconscious* mental processes and structures, a key difference between his psycho-analysis and other contemporary approaches to psychology, such as that of Wilhelm Wundt. Where Wundt claims that psychology “has to investigate the facts of consciousness, its combinations and relations, so that it may ultimately discover the laws which govern these thoughts and combinations” (Wundt 1912: 1),¹⁶⁷ Freud did not think that empirical psychology was limited to mental phenomena that were directly “observable” in inner sense. Rather, he suggested—or rather insisted—that the most important mental phenomena were *unconscious* and thus that empirical psychology would have to be *indirect*. Thus, for example, Freud argues that slips of the tongue and dreams “have a sense . . . , meaning, intention, [and] purpose” (Freud 1920/1963: 74)¹⁶⁸, but asks “To whom?” (Freud

¹⁶⁷ To be fair, Wundt admits the influence of “dark . . . fields of consciousness” (Wundt 1912: 74, 107, 109-10) and even the influence of past associations between ideas that might not be in consciousness at any given moment but can affect what does arise in consciousness. In that sense, Wundt anticipates the broad conception of the unconscious on which Freud would later focus.

¹⁶⁸ For elaborate discussion of slips of the tongue and similar phenomena, see Freud 1901/1960. For dreams, see Freud 1900/1965, 1920/1963(Part II), and Freud 1933/1964: ch. 29.

1920/1963: 267). That is, we can best interpret certain actions and conscious mental processes as expressing aims and purposes of which one is not conscious or even consciously rejects. But then the psychologist must posit *unconscious* aims and purposes operative in one's mental life. Today, even if relatively few psychologists are "Freudian" in any strict sense of the term, virtually all empirical psychologists agree that inner sense is an unreliable indicator of one's mental processes, that one must use indirect means to discern what is really going on in the mind of a human being, and that the analyst/experimenter can often know what is going on in a person's mind better than the person herself.¹⁶⁹

Freud was well aware of the importance of his emphasis on the unconscious: "the hypothesis of their being unconscious mental processes paves the way to a decisive new orientation in the world and science" (Freud 1920/1963: 26). Just as Kant's "Copernican" turn – discussed in chapter one – was one of his most important claims about the human being, Freud's most important and lasting impact on the question what is the human being comes from his own appropriation of Copernicus. As Freud puts it in his *Introductory Lectures on Psychoanalysis*:

In the course of two centuries, the naïve self-love of men has had to submit to two major blows at the hands of science. The first was when they learnt that our earth was not the center of the universe but only a tiny fragment of a cosmic system of scarcely imaginable vastness. This is associated in our minds with the name of Copernicus The second blow fell when biological research destroyed man's supposedly privileged place in creation and proved his descent from the animal kingdom and his ineradicable animal nature. This revaluation has been accomplished in our own days by Darwin But human megalomania will have suffered its third and most wounding blow from the psychological research of the present day which seeks to prove to the ego that it is not even master in its own house, but must content itself with scanty information of what is going on unconsciously in its mind. We psycho-analysts were not the first and not the only ones to utter this call to introspection; but it seems to be our fate to give it its most forcible expression and to support it with empirical material which affects every individual. Hence arises the general revolt against our science. (Freud 1920/1963: 352-3)

¹⁶⁹ For some classic examples of positing psychological processes of which one is not conscious, see Nisbett and Wilson 1977.

Like Kant, Freud compares his own radical revolution to that of Copernicus. But whereas Kant turns *back* towards placing humanity at the center of the universe, Freud – like Copernicus – continues a trajectory of undermining our privileged sense of self. Freud’s revolution, however, goes radically further than either Copernicus or Darwin because Freud challenges our centrality in *our own lives*. For Freud, much of who “I” am is not up to me. Not only am I not the center of the universe, nor even the pinnacle of life on earth; I am not even the most important force in my own soul.

As the preceding passage shows, Freud did not limit his conception of “unconscious” to rare and relatively unimportant mental processes of which one might not be aware at the moment. Freud shifts from “unconscious” as “the name of what is latent at the moment” to a whole theory of *the* unconscious as “a particular realm of the mind with its own wishful impulses, its own mode of expression, and its particular mental mechanisms” (Freud 1920/1963: 262). As his thought develops, he refines the structure of this realm of the mind, such that in his mature theory, the human mind can be seen as structured along two axes. On the one hand, there is a distinction between the conscious, the pre-conscious, and the unconscious. The “preconscious” is an “unconscious that is only latent and thus [that] easily becomes conscious” (Freud 1933/1964: 89, cf. Freud 1920/1963: 366-8), while the unconscious strictly speaking is further from consciousness. The “transformation” of what is truly unconscious to consciousness is difficult, requiring for its possibility something like psycho-analysis, and this transformation is always incomplete (Freud 1933/1964: 89). On the other hand, cutting across the divide between conscious and unconscious processes is a threefold distinction between the id (literally *das es*, or ‘the it’), the ego (literally *das ich*, or ‘the I’), and the super-ego.¹⁷⁰ The ego is what one normally considers one’s self, the generally-conscious, self-aware, decision-making regulator of one’s life, the “reason and good sense” that must “guide the powerful movement” of the id (Freud 1933/1964: 95-6). The super-ego performs the functions of “self-observation, conscience, and [maintaining] one’s ideals” (Freud 1933/1964: 83). What Kant calls the “predisposition to personality,” the conscience against which one measures one’s activities and evaluates them morally, is, for Freud, the “super-ego.”

¹⁷⁰ This account is different from Freud’s early theory in two important respects. In his early work, Freud had not yet distinguished between the ego and the super-ego, and he had identified the ego with consciousness and what he would later term the id with unconscious processes. For his most detailed discussions of this shift, see Freud 1923/1960 and Freud 1933/1964 (chapter 31).

The id is Freud's most important contribution to the theory of human nature: "The space [in the mind] occupied by the unconscious id [is] incomparably greater than that of the ego or the preconscious" (Freud 1933/1964: 98); in fact, *all* of the parts of the mind are really just modifications of the id, such that, for instance, the ego is "the better organized part of the id, with its face turned towards reality" (Freud 1933/1964: 116). The id is "the dark inaccessible part of our personality . . . , a chaos, a cauldron fully of seething excitations . . . [It] knows . . . no good and evil, no morality . . . [but only] instinctual cathexes seeking discharge" (Freud 1933/1964: 91). The ego is subject to a "reality principle" that requires that it moderate its desires so that they are consistent both with each other and with what is possibly achievable in the world. But in the id, "contrary [and unrealistic] impulses exist side by side" (Freud 1933/1964: 92). The law of non-contradiction and even "the philosophical theorem that space and time are necessary forms of our mental acts" do not apply to the id, where "no alteration . . . is produced by the passage of time." The fundamental principles of the id are the pleasure principle, which Freud associates pre-eminently with sexual desire, and a principle of self-destruction and aggression.¹⁷¹ The particular fixations and excitations of the id arise partly from innate natural instincts, partly from unconscious cultural inheritances, and largely from events – especially in infancy and childhood – that had a big effect on structuring one's sexual and aggressive desires but that one refuses, for various reasons, to admit to consciousness. Thus the famous Oedipal Complex arises from innate sexual desires in infancy that focus on the mother as a desire-object. Particular and forgotten details of one's infantile relationship with one's mother can then exercise powerful but unconscious influences on one's later life.

In relation to Kant, Freud's most important challenges relate to the general problem of self-knowledge and to the specific issue of the origin and nature of morality. With respect to self-knowledge, Freud is arguably even more modest than Kant's own claim that "we can never, even by the most strenuous self-examination, get entirely behind our own covert incentives" (4:407). But in Kant, this strong claim implied only a fairly straightforward humility about self-knowledge and was conjoined with what remained a fairly naïve approach to psychological investigation, one that privileged introspection and relatively straightforward inferences of motives from actions. Freud's complex psychic architecture, his willingness to posit unconscious forces radically at odds with what we experience in conscious life, and his development of a specific

¹⁷¹ See *Beyond the Pleasure Principle* (Freud 1920/1961) and *New Introductory Lectures* (Freud 1933/1964), especially chapters 32 and 34.

psychoanalytic methodology for unlocking the secrets of this unconscious all take him in a very different direction from Kant.

Far more important than his particular Copernican turn, however, is Freud's revaluation of the value of conscience.¹⁷² For Kant, what it ultimately means for a human being to reach his potential is for him to live autonomously, to live in accordance with the normative principles of thought, feeling, and action that arise from his participation in an intelligible world. In Freudian terms, Kant advocates a complete subordination of one's id (or "untamed passions," Freud 1933/1964: 95) and even one's ego to one's super-ego, one's conscience. But while Freud shares with Kant a commitment to articulate "rational explanations" for moral requirements rather than ascribing them to divine decree (Freud 1927/1961), he does not see the super-ego as the unconditionally good expression of human autonomy. Freud's "rational explanation" is ultimately in terms of the realistic satisfaction of desires rather than a defense of a categorically valid imperative governing humans as members of an intelligible realm. Even the super-ego, which arguably issues categorical imperatives, is seen by Freud as an internalization of one's infantile fear of punishment from one's parents and Oedipal desire to please one's mother (Freud 1933/1964: 77f.). It is a "vehicle of tradition" (Freud 1933/1964: 84) and largely responsible for the repressions and neuroses that haunt people in their adult lives. In sharp contrast to the moral-prudential goals of Kant's pragmatic anthropology, "the therapeutic efforts of psychoanalysis have [the] intention . . . to strengthen the ego, to make it more independent of the super-ego . . . so that it can appropriate fresh portions of the id" (Freud 1933/1964: 99).

While Freud's specific insights have waxed and waned in terms of their importance for contemporary psychological practice,¹⁷³ his legacy continues to influence contemporary thinking about the human being in at least four important ways. First, Freud brought the unconscious mind into center stage for psychological explanation of human thought and action. Today, even those who distance themselves from Freud's specific theories often continue to think of mental life on the model of competing (or cooperating) psychic forces of which one is only rarely conscious. Second, there continue to be vibrant psychoanalytic practices, and psychoanalytic techniques for diagnosing and treating mental disorders continue to be used in contemporary medical psychiatry. Third, near the

¹⁷² Arguably, this is a point Freud took from Nietzsche.

¹⁷³ Peter Kramer, for example, notes that at Harvard Medical School in the 1970s, "there was no distinction between studying psychiatry and following Freud" (Kramer 2006: 10), but many psychology textbooks today largely dismiss him.

end of his career, Freud applied his general models of explanation to historical and cultural analysis, where they continue to be widely appropriated by anthropologists and cultural critics. And finally and most importantly, Freud's Copernican Revolution continues to exercise a profound influence on general human self-conceptions. The sort of naïve assertions of self-awareness that dominated earlier attempts to know and express oneself have been replaced by suspicion of our own self-awareness and a willingness to accept that who we are is largely the product of psychological forces that we do not control and of which we are often unaware.

V. Conclusion.

The last book that Kant published during his lifetime, published just after his *Anthropology from a Pragmatic Point of View*, was entitled *The Conflict of the Faculties*. The book consists of three essays, taking on each of the three academic "faculties" – what we would now call "departments" or "disciplines" – against which philosophy, at the the end of Kant's life, was contending for relevance. At the end of the 18th century, these competitors were limited to jurisprudence, medicine, and, first and foremost, theology. The targets were misplaced, but the title of Kant's work remains apt. There is still a conflict of faculties for dominance in answering the question of what the human being is, with biology, psychology, sociology, history, "anthropology" (in the modern sense), literature, art, and literary theory taking the place of theology. For a brief period culminating in Kant and Hegel and in what Nietzsche aptly described as the "death of God," philosophy dominated cultural reflections on what it means to be human. But as theology receded, philosophy increasingly lost its influence to other arts and sciences.

This chapter has surveyed four of the most important figures who helped bring about this shift, and whose conceptions of human nature continue to be profoundly relevant and deeply influential today. In very different ways, Marx, Darwin, Nietzsche, and Freud helped open new vistas for understanding ourselves. Even within philosophy, many of the most important ways of addressing the question of human beings go back to one or more of these four thinkers. But each also helped shift Kant's question out of philosophy, into biology, psychology, other human sciences, and the arts. In the next chapter, we will take up the contemporary philosophical heirs of Darwin and Freud, who seek

answers in the sciences of biology and psychology. Chapter nine turns to thinkers who emphasize the contingency, historicity, and diversity of human natures, drawing from the historical and genealogical approaches of Marx and (especially) Nietzsche. And in chapter ten, we look at existentialist approaches to the self, approaches that in many respects trace themselves to Nietzsche's movement "beyond good and evil," towards a radically self-affirming conception of the (super-)human self.

Chapter 8: Scientific Naturalism

When Kant answered the question “What is the Human Being?,” biologists still took seriously the idea that every human being might literally and physically have “pre-existed” in Eve’s womb¹⁷⁴ and that species were eternal creations. Kant himself despaired of finding a “Newton . . . of a blade of grass” (5:400) and emphasized that scientists “do not know cranial nerves and fibers, nor do [they] know how to put them to use” (7:119). Empirical psychology was based more on introspection than strict scientific methodology and was not yet distinguished from philosophy.

Things have changed. Consider just a few highlights of our scientific knowledge about human beings:

- The Human Genome Project has successfully mapped humans’ genetic code and we increasingly understand both where we came from and how our genes direct our development
- MRI and CT scans have detailed the structure of the brain
- PET and fMRI scans can now track the brain activity of human beings involved in specific mental tasks
- Studies on non-human primates have shown possible origins of human altruism, language, culture, and even our sense of justice.
- Psychologists have developed models of unconscious motivation, and new methodologies (such as neural mapping, controlled correlational studies, and double-blind experiments) have begun to transform empirical psychology into a rigorous science.

In addition to these very general developments, scientists have recently made a number of counter-intuitive discoveries with the potential to dramatically change our sense of what it is to be human. For just a few examples,

- Benjamin Libet and others provide evidence suggesting that unconscious physical processes in the brain precede and cause conscious choices
- A biases and heuristics program in contemporary psychology offers evidence that irrationality is widespread and unrecognized even in the most careful and thoughtful human beings
- Situationist psychology provides evidence that much human behavior is determined by context rather than by character

¹⁷⁴ Perhaps even more dramatically, scientists like Anton Leeuwenhoek – the first scientist to observe sperm with a microscope – and Nicolaas Hartsoeker argued against this view with purported observations in microscopes of very tiny human forms seen within male spermatozoa.

These developments cover only a small fraction of the progress in human biology and psychology, without even mentioning the contributions of economics, sociology, anthropology and history to understanding human beings.

Taken together, this scientific progress not only calls into question fundamental aspects of Kant's anthropology but also offers some hope that the question "What is the human being?" can be answered by science rather than philosophy. In other words, scientific progress provides hope for scientific naturalism about human beings. Scientific naturalism is the view that everything that is real is part of nature, the world that is investigated by the natural sciences (including biology and scientific psychology). Generally, people tend to think that questions such as "What is the emu?" or "What is the monarch butterfly" or "What is oxygen?" are sufficiently answered, in principle at least, by fully developed scientific accounts of emus, butterflies, or oxygen. Many have suggested that human beings are not fundamentally any different, that the best answer to the question "What is the human being?" is just whatever our best biological and/or psychological theories say the human being is. Philosophy has nothing distinctive to contribute to understanding human beings; instead philosophy should simply "clarify and unify" what is given by science (Dennett 2003:15).

One task of this chapter is to explain and critique scientific naturalists' answers to Kant's question. Because Kant's way of dealing with science is not naturalist in this sense, this chapter also provides Kantian responses to recent scientific developments. Given the breadth of empirical research on human beings and the range of naturalist approaches to human beings, this chapter limits itself to discussing a few highlights from recent empirical research and its philosophical appropriation. By the end of my brief survey, I hope that simplistic reactions to naturalism will become more complicated. Naturalism does not flow as neatly from the progress of science as its proponents might have hoped, but it also has much more adequate resources for dealing with important aspects of our self-conception (such as freedom, creativity, and morals) than many of its opponents feared. Moreover, understanding human beings as natural beings provides valuable resources for actually helping us to be better human beings, but the value of science is greatest, I argue, when its insights are incorporated into a broadly Kantian anthropological framework.

Because scientific naturalism often involves a commitment to "materialism" – the view that there is nothing non-material (such as a soul) – and "reductionism" – the view that non-physical processes such

as cognition can be understood in terms of (or “reduced to”) physical processes, I start by looking at the most thoroughly materialist and reductionist approach to human beings: cognitive neuroscience, which investigates human reasoning, emotion, decision-making, and even creativity from the standpoint of physical processes in the brain. I then turn to evolutionary biology, which provides an account of how human beings have developed from more primitive biological ancestors. Such an account is necessary to complete the materialist naturalism of neuroscience, since without an account of the origin of the brain, one might posit – as some creationists do – that even if what it is to be human can be explained physically, the physical structure of the brain could not have come about through natural processes. Evolutionary biology also provides a scientific methodology for thinking about human beings that is not wholly dependent on neuroscience, and thus opens the door to a different sort of naturalist explanation of cognition, consciousness, culture, morals, and even freedom. Finally, I examine current trends in empirical psychology. Psychological naturalism is consistent with materialism but does not depend on it. One can hold that empirical psychology provides everything we need to explain the human mind without believing that psychological processes are reducible to physical processes. And philosophers have increasingly used psychological theories about human beings to develop naturalistic approaches to epistemology (what we can know) and ethics (what we should do).

I. Human Brains: Neuroscience and the Philosophy of Mind

In 1848, 50 years after the publication of Kant’s *Anthropology*, an accidental explosion sent an iron rod through the head of Phineas Gage, a railroad worker in Vermont. After recovering from the initial shock, Gage arose, rode into town awake and alert, and saw a doctor. Within two months, Gage was said to be cured, and by all indications was perfectly functional. But whereas Gage before the accident had been a polite, well-balanced, self-disciplined worker; Gage after the accident was “fitful, irreverent, indulging . . . in the grossest profanity . . . , manifesting but little deference for his fellows . . . , capricious and vacillating” (Valenstein 1986:90, cf. Damasio 1994). A physical alteration to Gage’s brain seems to have engendered a wholesale transformation in his character.

Gage's case is not unique; physical brain injuries have long caused mental and dispositional changes in human beings. And recent years have brought increasingly fine-tuned accounts of the parts of the brain responsible for different mental functions. At first, such scientific developments occurred primarily through careful analyses of victims of accidents like Gage's. But since the mid-1970s, PET and fMRI scans have made it possible to scan the brains of normally functioning adults performing different mental tasks. This brought neuroscience to a whole new level, resulting in increasingly fine-tuned maps of different control centers in the brain. Scientists have identified specific parts of the frontal and temporal lobes as loci of linguistic activity, a primary projection area in the parietal lobe that controls most motor activity, and C-fibers in peripheral nerves of the somatosensory area that are instrumental in feeling pain. At the same time, studies of neurons and glial cells at the cellular level help scientists understand brain activity and development.

While direct studies of brain activity have been an important source of greater understanding of the physiological bases of human mental life, other developments have provided analogies and models for thinking about the brain. Computers have been particularly important in two phases of thinking about the neurobiological basis of cognition. Even before functional computers existed, the idea of the brain as a computer was posited as a metaphor for thinking about human mental processes. The "Turing Machine," an early theoretical model for a sort of machine that could engage in basic "cognitive" tasks such as arithmetic and the construction of grammatically correct sentences, is now a commonplace metaphor for human cognition. In the early days of computing, the dominant model was to think of the brain as a sort of linearly processing computer. With the rise of computer networks as powerful technology, the idea of a "neural network" has taken off as a model for thinking about the brain. Just as computers can be networked together, neurons can be networked together to create an "everything-connected-to-everything" neural network capable of building connections based on past experience.¹⁷⁵ Even more recently, the use of "parallel processing" in computing – where multiple computers work on different parts of a process "in parallel" and then reassemble the results – has been used as a model for humans' "unconscious parallel processing (in which many inputs are processed at the same time, each by its own mini-processor)" (Pinker 1997: 140).

Applying these neuroscientific discoveries to thinking about the human mind and its relationship with the brain has become a central

¹⁷⁵ For a helpful discussion relating neural networks to computer networks, see Pinker 1997: 99-131.

problem within the subfield of philosophy called “philosophy of mind.” One view of the mind, which might seem to be the most intuitive implication of the close correlation between brain-states and mental-states, is “eliminativist materialism” about mental properties, the “identification of mental states with physical states,” such that what seem to be mental states are really physical states. Paul Churchland compares the case with that of color: “In discriminating red from blue . . . our external senses are actually discriminating between subtle differences in intricate electromagnetic . . . properties of physical objects . . . The same is presumably true of our ‘inner’ sense: introspection” (Churchland 1984: 29). Such a view represents a strong scientific naturalism, in that there is nothing more to human beings than our (neuro)physiology.¹⁷⁶ It also implies materialism and reductionism: what seems mental is really physical, and psychology is wholly reducible to neurobiology. As Daniel Dennett has put it, “there is only one sort of stuff, namely matter – the physical stuff of physics, chemistry, and physiology – and the mind is somehow nothing but a physical phenomenon. In short, the mind is the brain” (Dennett 1991: 33).

There are some important problems with eliminativism, however, that have led philosophers of mind to articulate alternatives. Three of the most important problems are qualia, multiple-realizability, and intentionality/normativity. The term “qualia” refers to the subjective feel of particular mental states. As Thomas Nagel puts it in his famous essay, “What is it like to be a Bat?”, the subjective character of an organism’s mental states entails that “there is something that it is like to be that organism-- something it is like for the organism” (Nagel 1974:476).¹⁷⁷ Many philosophers have come to think that it is precisely this subjective character of our mental states that makes the mind irreducible to the brain states investigated by neurobiology. The problem of multiple realizability arises for many attempts at reductionism, including the reduction of the mind to the brain. In its most basic form, the problem is that phenomena that appear at one level of explanation are realizable in many different ways at a different level of explanation. Pain, for example, might be instantiated in many different neurobiological configurations. And even if these all share a common element in humans (such as the firing of C-fibers), one might find other animals (and one could certainly

¹⁷⁶ Obviously no one would deny that, in some sense, our physiology more broadly is part of what it is for us to be human. But contemporary philosophers and scientists, like Kant, typically emphasize the human mind as particularly distinctive to human nature. That said, there has been an increased attention in recent years to the way in which the “mind” may be located, not in the brain per se, but in the body as a whole or even beyond the body in the world in which we live (see Dennett 2003 and Noe 2009.)

¹⁷⁷ For a helpful overview (with bibliography) of recent discussions of qualia, see Tye 2007.

imagine other creatures) that feel pain with a different neurobiological architecture (with some other neural element playing the role of C-fibers) (see Putnam 1967). Moreover, there may well be psychological laws that cannot be formulated in physical terms, because of the different ways in which psychological states can be realized. A simple psychological law (like “fear of a lion provokes a fight or flight response”) might be untranslatable into strictly physical terms since the physical states associated with a particular instance of fear might not fit into a physically delineable type that would be consistently correlated with a physically delineable type of effect corresponding to “fight or flight.” Insofar as psychological laws are both informative and untranslatable, eliminativism fails to capture the whole truth about human mental life. The problems of intentionality and normativity come from the fact that many human mental states seem to about something and/or have the potential of being right or wrong. One is not merely afraid, but afraid of a lion. One does not merely have a belief-state, one believes (rightly or wrongly) that the lion is going to attack you. One does not merely have a volitional state, one decides (rightly or wrongly) to run away from the lion. In each case, one seems to simply have a brain-state. And while a brain state can be caused by something else (say, the perception of a lion), it is not clear how a brain state can be of something else, nor how a brain-state could be true or false or right or wrong (it just is what it is).¹⁷⁸

These three considerations have led many philosophers of mind to develop alternatives to eliminative materialism about the mind. One alternative is Descartes’s substance-dualism. Descartes was well aware of the close connections between mental states and the brain¹⁷⁹ but saw mental changes as irreducible to physical brain-changes. Instead, Descartes described the mind-brain connection as a mutual influence between two distinct substances: a non-material soul, or mind, and a material part of the brain. The soul experiences qualia and engages in intentional, normatively-governable mental activity. The body is purely material and acts on other material things. These two substances are capable of interaction, so that changes in one can cause changes in the other, but neither is reducible to the other and either could in principle persist without the other.

¹⁷⁸ These examples differ in important respects. Intentionality is not identical with normativity, and different forms of normativity (epistemic, prudential, moral) are not identical with one another. But intentionality and the normative dimensions of mental states all raise a common problem for eliminativism, which is that we ascribe properties to mental states that seem untranslatable, even in principle, into anything that could be said of a brain-state.

¹⁷⁹ He even famously located the “pineal gland” in the brain as the locus of higher cognitions in human beings.

Currently, most philosophers of mind reject both substance-dualism and eliminativism in favor of a more complex view that can be called functionalist property-dualism with token-identity between mental and physical states (Botterill and Carruthers 1999). Each element of this description is important. Property-dualism is way of responding to the problems with crude materialism without falling into a full-blown substance-dualism. The idea is that there are two irreducibly distinct sorts of properties of human beings, our physical properties and our mental properties. These are not different substances, but they are irreducible to one another, such that one could make true claims about the mind – say, claims about qualia or connections between mental states – that could not be translated into claims about the brain. Functionalism is a way of making sense of what one refers to when one describes a particular type of mental state: “functionalists characterize mental states in terms of their [functions or] causal roles, particularly, in terms of the causal relations to sensory stimulations, behavioral outputs, and other mental states” (Block 1980:172, cf. Putnam 1967). And token-identity is the view that each particular mental state “token” – such as the initial feeling of pleasure I experienced last night as I began eating dessert – is identical to a particular brain-state “token.” This provides for an important measure of materialism, since each individual mental state is identical to a particular physical brain-state, but it avoids problem of multiple realizability, since each type of mental state can be realized in many different types of brain-state (see Davidson 1970).

To some, Kant’s view on the relationship between mental states and brain-states might seem similar to Descartes’s substance-dualism. Kant distinguishes between the noumenal thing-in-itself and the phenomenal appearances, and Kant specifically applies this distinction to human beings, who are both transcendently free things-in-themselves and embodied, empirical appearances. But this apparent parallel with Cartesian dualism is misleading. While Kant makes use of the distinction between transcendental and empirical anthropology to make sense of some of the problems that lead philosophers of mind towards various dualisms, his own account of Cartesian dualism locates this dualism within the realm of appearances (see e.g. 28:680-1). In fact, since the category of “substance” is a category that structures the empirical world, Kant’s distinction between things-in-themselves and appearances cannot, except in an analogical sense, be considered a “substance-dualism” at all. For Kant, the “mind” is an empirical object available to inner sense, and Kant must therefore ask to what extent this empirical mind is reducible to something purely physical. Kant thus distinguishes empirical-substance-dualism, by virtue of which the mind

and body would be empirically distinct substances, from transcendental-dualism, according to which the mind-in-itself must be distinguished from the empirically-knowable-mind.¹⁸⁰

Kant is certainly committed to a transcendental-idealist-dualism that implies two irreducible perspectives on mental life. Transcendental anthropology is distinct from empirical anthropology, and insofar as there is an empirical mind, it can be distinguished from its noumenal ground. But it is far from obvious that substance dualism is needed to preserve both a standpoint on the mind-in-itself that is irreducible to empirical descriptions and the possibility of normative claims about human thoughts, feelings, and choices. Even if some metaphysics of mind is needed to ground this distinction between standpoints – something about which contemporary interpreters of Kant sharply disagree – one could simply draw on a sort of transcendental property dualism according to which the human mind has properties “in-itself” that are irreducible to its empirical properties.¹⁸¹ Kant’s transcendental idealism thus commits him to some sort of dualism, but not to a distinction between two interacting substances. And in this way, Kant actually provides a way in which one can be a materialist about the *empirical* mind while reserving a space for normativity and other “from-within” aspects of the mind understood transcendentially.

Kant’s transcendental idealism also does not commit him to any empirical dualism. Within the realm of appearances, Kant could accept a strictly eliminativist philosophy of mind without threatening his transcendental anthropology. Even Kant’s empirical anthropology could be preserved on an eliminativist reading. One would simply need to translate the psychological laws that Kant lays out there into physical

¹⁸⁰ For discussions of Kant’s philosophy of mind, see Ameriks 1982a, Aquila 1983, and Brook 1994. There could be yet a third dualism, which we could call transcendental-ground-dualism, by virtue of which the transcendental ground of the mind is distinct from the transcendental ground of the body. If there is any argument for transcendental-ground-dualism, it would be a moral one. Since we hold people responsible for their choices, we can identify a free noumenal ground for those choices. Since we do not hold them responsible for their physical states in general, we might think that the noumenal ground of physical states for which we are not morally responsible is different than the ground of states for which we are morally responsible. Kant does not offer this argument, and his transcendental anthropology need not commit him to this distinction, but it opens up room for a new, morally-grounded, way of thinking about something like a mind-body dualism. (For Kant, we are never responsible for bodily states themselves, but only for certain higher volitions that might have bodily conditions or effects. There are, however, lots of mental states for which one is not directly responsible – including most perceptions, cognitions, and emotions – so the sphere of the “body” would be much wider than what we normally think of as our physical body.)

¹⁸¹ For a discussion of different approaches to the metaphysics of the transcendental-empirical distinction, see chapter one and Ameriks 1982b. It is noteworthy that even as Ameriks raises serious problems for non-metaphysical readings of Kant’s transcendental idealism, he tacitly endorses a “two-aspect” rather than “two-substance” reading of human agents.

laws of the brain. Kant's claim that "feelings depend on cognitions" would become a claim about the dependence of certain brain-states upon others. And underlying natural predispositions would be reducible to structural limitations on the physical operations of the brain.

Nonetheless, Kant rejects eliminativism for two main reasons. First, because "the soul can perceive itself only through the inner sense" (12:35) and inner sense is purely temporal while the physical body is always spatiotemporal, the most that physiological explanation could ever do it to explain "the matter that makes possible" mental phenomena (12:35). Mental phenomena as such have a character that is irreducible to the physical. The content of an inner experience – a feeling of fear, for example – thus cannot be identical to the content of an observed brain-state. Second, just as Kant argues that empirical psychology must posit multiple different kinds of mental state to make sense of the phenomena of human mental life, he argues that science in general cannot depend upon purely physical causes in making sense of the behavior of living (and especially human) things. In general, for Kant, science should use as few general principles as possible, but as many as are truly needed to make sense of observed phenomena. Just as Newton legitimately (according to Kant) positing gravitational force in order to better model physical motions, Kant posits "preformed" teleological and psychological predispositions to better explain living and animal behavior. And just as Newton did not reduce gravitational force to the mechanistic forces of inertia and collision that dominated 17th century physics, Kant does not reduce psychological forces to purely physical ones.

Neither of these Kantian arguments need imply empirical-substance-dualism, however. The first argument – based on the distinction between inner and outer sense – is a sort of qualia argument, put in terms of Kant's general account of the difference between the way inner and outer states appear to human knowers. Inner states have a certain feel – non-spatiality – that outer states necessarily lack. But this lack of equivalence does not imply any difference of substance between mind and body. In the same way that the irreducibility of auditory to visual sensations is consistent with having both kinds of sensations of the same object, perceptions of mind in inner sense and of brain in outer sense could be irreducibly distinct perceptions of the same thing. Kant even does some speculative neuroscience, suggesting chemical processes in "the water of the brain" that might underlie the processes of "separating and combining given sensory representations" (12:34). The second argument – the need for purely psychological laws – is an empirically-contingent one that might be falsified given neuroscientific progress. At present, however, optimism that all psychological laws will

eventually be translatable into neurophysiological laws is merely a scientific ideal; and the multiple-realizability of mental states provides reasons for thinking that even the most sophisticated neuroscience will still leave room for properly psychological laws in explaining human thoughts and actions. But this argument, too, does not require a substance-dualism, only an irreducibility of the relevant laws, or, for Kant, powers. The same substance can have different powers – as Kant clearly thinks is true of the human soul – and there is no reason that the physical powers of the brain and the mental powers of the mind could not be distinct powers of the same thing (the brain-mind).

In general, then, Kant's anthropology puts him in an excellent position vis-a-vis contemporary debates in the philosophy of mind. Kant's argument based on the non-spatial character of inner sense contributes an important variation on the qualia argument for the difference between mind and body. His generally Newtonian approach to science provides a basis for distinguishing psychological and physical laws, one that is appropriately modest about the prospects for neuroscience, not limiting these prospects a priori but also recognizing the still-present need for non-physical laws to fully make sense of human (and other living) beings. In both respects, Kant's philosophy of mind anticipates some of the most important contemporary arguments for an empirical dualism between mind and body. Kant's transcendental idealism, wherein the mind as seen from-within and bound to normative laws is distinguished from the mind as an object of empirical knowledge, further enriches his philosophy of mind. Moreover (as I argue in more detail below), Kant rightly shows that the distinction between the empirical mind as the object of psychology and the empirical body as object of biology is insufficient to account for normativity. With respect to normativity, the problem is not eliminativism or materialism but any form of naturalism. Treating the mind as an object of description according to natural laws is insufficient for giving an account of the mind as bound by normative laws. The normativity problem calls for a different sort of solution than the problems of irreducibility and qualia.

Kant opposes eliminativism from two directions, neither of which requires a commitment to full-blown Cartesian dualism. The non-spatiality of inner sense and the (so far) irreducibility of psychological laws to physics ones give good reasons to distinguish mental properties from physical ones in empirical descriptions of human beings. Transcendentally, the normativity of the from-within standpoint on human mental life requires distinguishing this standpoint from *any* empirical standpoint (whether psychological or physical). Kant's distinction between transcendental and empirical anthropology both

allows for these necessary distinctions and provides a natural way to incorporate neuroscientific insights into his overall philosophy without compromising his transcendental philosophy.

Even if Kant's overall account of the human being is compatible with general developments in neuroscience, particular neuroscientific findings challenge Kant's particular claims about human beings. Most of these findings require only minor modifications of or additions to Kant's empirical account of human beings, but some recent research suggests pictures of the human mind that seem to challenge some of our (and Kant's) most fundamental conceptions of what it means to be human. A study by Benjamin Libet, for example, has subjects flick their wrists while researchers scan their brain activity with an EEG. Subjects flicked their wrists at will, and Libet found that each wrist-flicking was preceded by a consistent EEG pattern. Libet then asked subjects to look at a simple, rapidly-moving clock-face and note the position of a dot (equivalent to a clock-hand) at the moment they made the conscious decision to flick their wrist. The surprising result is that the EEG pattern that brings about wrist-flicking preceded the conscious decision to flick. As Libet puts the results of his study, "The initiation of the freely voluntary act appears to begin in the brain unconsciously, well before the person consciously knows he wants to act! Is there, then, any role for conscious will in the performance of this voluntary act?" (Libet 1999:51). Or, as the same result was put by Dennis Overbye for the *New York Times*, "The decision to act was an illusion, the monkey making up a story about what the tiger had already done" (Overbye 2007).

Against such apparently paradoxical conclusions, Kant's humility about science reduces the threat of Libet's findings without requiring bizarre accommodations. When Libet tries to answer his question of whether there is "any role for conscious will in the performance of a voluntary act," the most he can do is to give the will a sort of "veto" over the flick based on the fact that "it must be recognized that conscious will does appear about 150 milliseconds before the muscle is activated" (Libet 1999:51). This attempt to salvage some remnant of freedom is, of course, implausible. The fact that the neural process is not yet finished by the time one is consciously aware of one's decision to flick does not imply that consciousness can bring that neural process to a halt.¹⁸² For Kant,

¹⁸² Daniel Dennett (see Dennett 2003) has, somewhat more helpfully, pointed out a variety of mental architectures that could explain how one might mistake the location of the dot at the time of one's decision and thus report a time that was later than one's conscious decision-making. The connection between events in inner sense and events in outer sense – the brain – is likely to be complicated. There is no reason to think

however, “conscious decision-making” is an ambiguous phrase, one that can refer to either an object of inner sense – one’s introspection of a particular event of cognition giving rise to a volition – or to a transcendental perspective on action, the standpoint of considering alternatives or evaluating choices “from-within” for the purposes of deliberation or the ascription of moral responsibility. Empirically, Libet’s experiment need not raise any red flags, since Kant’s empirical dualism is consistent with conscious psychological states being correlated with and even caused by physical states. Explanation in terms of conscious decisions takes place at a different level than explanation in terms of brain-state fluctuations, so what matters in this case is merely that conscious decision takes place, not the timing of that decision relative to the physical changes that underlie it.

The greater threat might seem to be to the transcendental perspective on action, since Libet’s experiment makes it look as though brain-states must be the causes of choices rather than vice versa, since they precede those choices. But for Kant, the priority of free choice over the determinism of the empirical world is not temporal. The suggestion that brain-patterns precede conscious choices seems threatening because we assume that unless our choices come temporally first and determine the structure of the world, we cannot really be responsible for them. This is just the conception of freedom that Kant’s transcendental anthropology rejects, by showing that we can be responsible for actions even if, from a scientific perspective, we need to see those actions as the results of prior causes in a deterministic world. It should come as no surprise that scientists looking for causes of human actions eventually find them, since they modify their overarching theories in order to make human behavior fit into the same causal-determinist models as other phenomena. But the fact that scientists can and must continue to refine their theories to develop better and better causal models of human behavior does not change the nature of our transcendental standpoint on human action. From-within the standpoint of a deliberating agent, we must still see our actions as the free results of choices that *ground* those actions and are undetermined by physical causes.

that the moment at which one becomes aware of a decision is identical to the moment at which one makes the decision. Even empirically, choosing is one thing and introspectively reflecting on that choice is another. Thus the fact that one’s consciousness of one’s choice post-dates the physical mechanisms that correlate with that choice may be interesting, but it is not particularly threatening to Kant’s conception of the human being. Dennett’s argument is one with which Kant could entirely agree, but for Kant, such theorizing about possible looseness in the empirical account is not necessary. In principle, for Kant, freedom will never be found in the timing of events in the brain, nor could any particular mental architecture be more undermining of freedom than any other. All human choices are, when seen empirically, the determined results of prior natural causes.

Before leaving this section, it is worth noting one further context for thinking about Kant's relationship with contemporary neuroscience. Neuroscience affects most people's lives neither through knowledge of particular scientific theories about brain-states nor through philosophical reflection on the nature of mind, but through psychopharmaceuticals used to improve psychological health. To some extent, Kant would be pleasantly surprised by this use of neurobiology for improving human lives. Although he refers to "inquiries as to the manner in which bodily organs are connected with thought" as "eternally futile" (10:146), Kant is willing and even eager to appeal to physiological treatments when they are reliable and available (see 7:213, 220). And when Kant objects to physiological approaches to pragmatic anthropology, he refers to approaches that emphasize bodily bases of mental states and that therefore cannot be put to any practical use. For Kant, a physiological focus implies practical uselessness because of the limited knowledge of how to manipulate the body to bring about shifts in mental states.

He who ponders natural phenomena, for example, what the causes of the faculty of memory may rest on, can speculate back and forth . . . over the traces of impressions remaining in the brain, but in doing so he must admit that in this play of his representations he is a mere observer and must let nature run its course, *for he does not know the cranial nerves and fibers, nor does he understand how to put them to use for his purposes*. Therefore all theoretical speculation about this is a pure waste of time. (7:119, emphasis added)

The improvement of neuroscience has the potential to transform a formerly useless physiological anthropology into an important part of a genuinely pragmatic anthropology.

But Kant's pragmatic anthropology, while it would certainly appropriate contemporary neuroscience for practical purposes, also provides an important counterweight to and caution about the clinical approaches that had already begun in the 18th century and have developed even further today. For Kant, "Medical science is philosophical when the sheer power of man's reason to master his sensuous feelings by a self-imposed principle determines his manner of living" (7:101). Kant's concern with physiological approaches to mental disorder is not *merely* that they do not work. Such approaches also put one's mental life in the hands of someone else. Rather than taking charge of one's own mental well-being, one "has a doctor who decides upon a regimen for me" (8:35).

And this turning over of one's own mental capacities to another grates against the autonomy that Kant repeatedly emphasizes in both morals (see especially *Groundwork* and *Critique of Practical Reason*) and intellectual life (see "What is Enlightenment?").¹⁸³ The increasing dependence on pharmaceuticals can also encourage people to abdicate personal responsibility for failings that, however physiologically influenced, are nonetheless expressions of a character that is ultimately free. Kant rightly notes that empirical anthropology is most important not as a theoretical endeavor but as a part of a practical discipline oriented towards improving human lives, and in that sense psychiatry is extremely valuable. But Kant also provides an alternative model of practical self-control that, while still being empirically-grounded, allows for genuine self-improvement rather than an abdication to others of the autonomy that is so important to being human. In the end, the overall structure of Kant's anthropology provides a framework for incorporating but also recognizing the limits of neuroscience in both practical life and theoretical self-understanding.

II. Humanity Evolves: Darwinism and the Fate of Humanity

However interesting the connection between the human mind and the human brain, contemporary neuroscience invites the question of how a physical system such as the human brain arose. Without a naturalist account of its origin, the human brain might seem literally miraculous, reflecting some supernatural design (much as computers reflect their human designers). As we saw in chapter two, Kant rejected attempts by his contemporaries to offer naturalistic explanations of basic human predispositions (8:110). But Darwin's theory of evolution by natural selection in the middle of the 19th century offered a major new theoretical framework for answering such questions. The 20th century saw a "Darwinian synthesis" between Darwin's theory of natural selection and Gregor Mendel's theory of heredity through the recognition that the random variations that Darwin left unexplained as brute inputs to his system could be understood as mutations of sub-cellular "genes" that were both heritable and susceptible to environmentally-induced mutations. The discovery of DNA by Watson and Crick in 1943 and the subsequent development of molecular genetics further explained the physical bases of genetic variation.

¹⁸³ As we will see in section three below, this is true even for non-neurological ways in which contemporary psychology encourages people to put their own mental and moral lives in the hands of others.

The immediate implications of the current biological synthesis between Darwinian natural selection and molecular biology for thinking about human beings are fairly straightforward. Like all life on earth, humans evolved from simpler organisms. Early in the history of our planet, molecules emerged that were capable of replicating themselves with slight variations. Those variations better at self-replication and persistence in the environment increased in number, and at a certain point reached levels of complexity that could warrant ascribing the label “life” to them. These self-replicating “organisms” competed for energy and other resources and, through natural selection, those better at replicating in their environments grew in number. The features that distinguish human beings from other animals are features that arose by means of molecular (primarily genetic) mutations that were preserved through this process of natural selection, whereby variations that add “fitness” – that is, allow survival and reproduction in greater numbers – grow more prevalent in the population. Human animals are well-adapted to our environments because earlier members of our genus that were not well adapted died and left no offspring. The human brain has the complex structure that it does because this advanced brain allowed ancestral humans to outcompete their closest relatives.

As a tool for understanding our world and ourselves, evolutionary biology has proven incredibly powerful over the past 50 years. Our knowledge of DNA allows us to more accurately diagnose genetic diseases and determine the likelihood that a disease will be passed on to one’s progeny. The success of the Human Genome Project in mapping the entirety of the human genome has helped us to identify genes associated with muscle disease, blindness, and deafness, and to understand the complex DNA sequences at the root of cardiovascular disease, arthritis, diabetes, and various kinds of cancer. The project has fueled new research with the aim of creating more specific and targeted treatments for many diseases, with fewer harmful side effects. Understanding natural selection and patterns of adaptation is central to developing and properly managing the use of antibiotics and vaccines, and evolutionary biology is at the core of our attempts to preserve endangered species and fragile ecosystems. Modeling lines of descent through genetic mapping provides guides population patterns over time and explanations for physiological differences across different populations, and DNA tests are used in forensics, parental rights, and tracing family genealogies.

Evolutionary models of human beings also provide a naturalist framework and scientific discipline to the popular philosophical pastime of armchair theorizing about human nature. Daniel Dennett, one of the foremost philosophical popularizers of Darwinism, explains,

Speculative exercises in agent-design have been a staple of philosophers since Plato's Republic. What the evolutionary perspective adds is a fairly systematic way to keep the exercises naturalistic (so we don't end up designing an angel or perpetual motion machine). (Dennett 2003: 217-8)

Rather than introspection or hand-waving about human nature, evolutionary theory forces one to explain, for any proposed physical or psychological feature, what effects such a feature would have on the fitness of an organism that possessed it. One cannot simply say that humans have features that would be nice to have or that would help explain particular behaviors. One must also give some account of how those features could have evolved through processes of natural selection. Given much recent work in applying evolutionary theory to human beings, there is reason to think that such accounts will prove to be more illuminating than one might have expected. One might even think that given the advanced state of human sciences such as evolutionary biology, there is little left for philosophy but to clarify and systematize "investigations in the natural sciences" (Dennett 2003:15).

Still, one might fear that evolutionary accounts would have difficulties making sense of central aspects of human life. Kant famously offers a sort of anti-evolutionary argument in the opening of his Groundwork, explaining that reason can have as its purpose neither human happiness nor reproductive success, since it is notoriously bad at promoting those, and certainly much worse than animals' instincts (4:395-6). Many also worry that any evolutionary approach to human beings will result in a picture of humans as hopelessly selfish animals seeking only to thrive and reproduce in a cut-throat world where "only the fittest survive." More generally, one might wonder whether central human concerns – morality, art, and even the sciences themselves, not to mention true love and religious experience – can be accounted for by evolutionary theory. Some of these concerns are concerns about naturalism more generally, questions about whether any theory that treats human beings as natural beings can accommodate central aspects of who we are. But some are tied to the particular kind of naturalist explanations offered by evolutionary theory.

In last 30 years – starting with the publication of E.O. Wilson's Sociobiology and Richard Dawkins' The Selfish Gene in the mid-70's – sophisticated Darwinian accounts of human nature have emerged that move beyond caricatures of evolution as implying that human beings are fundamentally nothing more than clever, selfish primates. The richness of these accounts is impossible to convey in this short chapter, but three

central issues – the evolution of altruism, the role of “memes” in evolution, and the nature of human freedom – give a sense of how evolutionary theory is used to make sense of aspects of human beings that might seem to transcend simplistic inferences from our descent by means of natural selection.

Evolutionary theorizing about altruism might seem oxymoronic because many see the claim that everyone is out for themselves as a central premise of evolution. But Darwin’s own *Descent of Man* emphasized that human fitness is enhanced through the development of “social instincts” that “lead an animal to take pleasure in the society of its fellows, to feel a certain amount of sympathy for them, and to perform various services for them” (Darwin 1981: 72). As our understanding of the processes of evolution grows, we see cooperative forms of natural selection playing key roles in the development of virtually all life on earth. A first approximation to altruism is present even in the most basic units of life on earth. The first few billion years of life on earth were dominated by “prokaryotes,” simple, single cells that “did everything for themselves.” If these cells moved, they moved themselves. If they generated energy, they generated it for themselves. If they broke down other cells to get organic materials for themselves, they had within themselves the resources to break down those cells. But about a billion years ago, some prokaryotes found themselves teamed up with others (by being incorporated into others without being broken down), and in some cases, these teams outperformed independent prokaryotes around them. These so-called “eukaryotes” prospered, and virtually all life today consists of complex cells that include “parts” that are descendents of these paired simple cells. Human cells contain, for example, mitochondria, which do most of the energy-processing in our cells and which have their own DNA,¹⁸⁴ and the vast majority (more than 99%!) of the genes in our bodies are in “non-human” micro-organisms living in our guts (Gill et. al. 2006).¹⁸⁵ Evolutionary “fitness” is not merely – nor primarily – a matter of killing off opponents. It can just as easily be a matter of cooperating in particularly effectively ways (cf. Dennett 1995). In the human case, for example, societies whose members cooperate tend to outperform those that are constantly at each other’s throats, so social affection and cooperation evolved among human beings.

¹⁸⁴ One can even trace maternal descent through these mitochondrial DNA because, unlike humans’ nucleic DNA, mitochondrial DNA comes entirely from one’s mother.

¹⁸⁵ Gill goes so far as to say that “humans are superorganisms whose metabolism represents an amalgamation of microbial and human attributes” (Gill et. al. 2006: 1355).

This importance of cooperation-as-evolutionary-fitness is made clearer by thinking about so-called “Prisoner’s Dilemma” scenarios. The standard Prisoner’s Dilemma is the following. Two suspects are being questioned for a crime. Each faces the following possibilities: If you betray your partner and your partner stays faithful to you, you go free and your partner gets life in prison. If you do not betray and your partner does, you get life and he goes free. If neither betrays the other, you’ll both be convicted only of minor charges for which you’ll spend only a short time in prison (say, a year and a day). If both betray, you’ll both probably end up spending a moderate amount of time in prison (say, 5-7 years). The ideal scenario overall is for both you and your partner to hold fast and say nothing; then you will both spend only a year in prison. But you both have strong incentives to betray, since whatever your partner does, you end up better off if you betray. The specific example highlights a general kind of case, one where the group as a whole (here you and your partner) would be better off if everyone adopted a particular course of action, but where each member of the group has an incentive to adopt a different, less optimal, course of action. For evolutionary theory, the Prisoner’s Dilemma might seem, at first blush, to pose a particularly pessimistic, even tragic, picture of life on earth. If evolution proceeds through a model of the “survival of the fittest,” then it looks like only betrayers will survive. Faithful humans who pursue strategies good for the whole will end up being evolutionary suckers, exploited by the betrayers looking out only for number one.

In fact, however, a sophisticated understanding of evolution shows that the Prisoner’s Dilemma has precisely the opposite implication: humans are more likely to evolve strategies that favor the group to which they belong than strategies that narrowly favor themselves, for several reasons. For one, situations like the Prisoner’s Dilemma described above are rare. Far more common are iterated Prisoner’s Dilemmas, in which one finds oneself in Prisoner’s Dilemma scenarios with the same people, or members of the same community, again and again. And in these contexts, selfishness (betrayal) is generally not the best strategy. In these cases, the best strategy tends to be some form of altruistic tit-for-tat. One starts by not betraying and continues to remain true to others, unless they betray (or are known to betray), in which case one prudently (or vindictively) betrays them in turn. Organisms – including humans – who are involved in iterated prisoner’s dilemma scenarios tend to thrive and reproduce most when they are altruistic but temper their altruism with what we might call justice and prudence. Human life is full of such scenarios, where cooperation is beneficial but a potential for exploitation exists. Thus even if evolution were simply a matter of the fittest

individuals surviving and reproducing, altruism tempered by prudence and justice will tend to evolve in human beings, since this configuration of dispositions is in the best interest of those who have them.¹⁸⁶

But the evolution of altruism is further enhanced by the fact that evolution is *not* simply a matter of fittest individual organisms surviving and reproducing. In the *The Descent of Man*, Darwin explains how “social qualities” such as “sympathy, fidelity, and courage” evolved “through natural selection.”

When two tribes of primeval man, living in the same country, came into competition, if the one tribe included . . . a greater number of . . . sympathetic and faithful members, who were always ready to warn each other of danger, to aid and defend each other, this tribe would without doubt succeed best and conquer the other. (Darwin 1981: 162)

Even if, among individuals, a more selfish individual were likely to have more offspring and thereby take over that population, one can also adopt a higher standpoint from which one sees competition amongst populations. To go back to our initial Prisoner’s Dilemma case, if one compares individual criminals, ones that betray their confederates will spend less time in prison than those that do not. But if one compares different criminal gangs, then the gangs whose members steadfastly refuse to betray confederates will spend less time in prison than those that are full of betrayers. Gangs that develop a tendency not to betray will be more successful than those that do not. In the evolutionary context, mutations that make members of society altruistic (at least vis-a-vis other members of their own society) will make that society as a whole more fit than other human populations. More altruistic societies (that is, societies whose members are altruistic) will tend to win out over less altruistic ones, and altruism will gradually become part of human nature.

¹⁸⁶ One might wonder whether it wouldn’t be better to have a more sophisticated strategy, whereby one cooperates only when one will be found out for betraying, but betrays whenever one can do so secretly, or whenever one will not need to depend upon those who become of one’s secrecy. There is evidence that such a strategy can be somewhat successful, but there are two problems for it. First, it requires a lot of time and effort to assess one’s situations in this fine-grained a way. The result is that it is probably more efficient simply to forego possible opportunities for safe exploitation rather than suffer either the cost of ensuring that one has diagnosed the scenario properly or the risks of getting it wrong. Second, the evolution of such a fine-grained strategy of deception will provoke counter-strategies of detection. There is good evidence that in human beings, whose brains provide the cognitive power to assess situations in fine-grained ways, both deception and counter-deceptions strategies have evolved. The result is a human nature that includes altruism and a sense of justice alongside selfish and deceitful tendencies and strategies of prudence for dealing with other selfish, deceitful people.

This point about group selection can also be made from the opposite perspective. Evolution works on groups as well as individuals, but especially in the light of the synthesis between Darwin and molecular genetics, biologists typically emphasize the primary locus of evolution as the gene, rather than the individual organism or group. In *The Selfish Gene* (1974), Richard Dawkins helpfully highlights that evolutionary selfishness takes place at the level of the genes. This does not mean that there is a gene “for selfishness,” but rather that genes code for what best allows the gene to survive and replicate (over the long term). And this gene-centered point of view has the same effects as the group point of view. The notion of “kin-selection” is a case in point. The idea behind kin selection is that because members of the same family share much of their genetic material in common, genes will tend to survive and propagate insofar as they give rise to instincts to protect one’s kin. The “interest” of one’s genes might well require that one sacrifice oneself for the sake of a sibling or child who shares copies of those genes. Insofar as human behavior is genetic (instinctual), it will tend at least to be altruistic towards groups to which one is genetic similar. Importantly, for the human individual, there is nothing “selfish” about this behavior; one’s genes code for behavior that – from the standpoint of the individual – is *genuinely* altruistic. In the end, whether one looks at evolution from the standpoint of individuals involved in iterated prisoner’s dilemmas, or groups competing for fitness, or selfish genes striving to replicate themselves, there is good reason to think that evolution by natural selection will give rise to human beings (and other organisms) that are cooperative and altruistic in the main senses in which those concepts are important to our self-conception.¹⁸⁷

Even if evolutionary theory is not committed to a conception of human beings as thoroughly selfish, however, it may seem ill equipped to account for the great cultural achievements of human beings. Can Darwinian evolution explain how we came to construct religious institutions, or create great works of literature, or develop complex societies and governmental systems and economies, or acquire scientific knowledge? In recent years, evolutionary theorists have developed various theoretical tools for making sense of these tendencies in Darwinian terms. Among these, arguably the most important is the concept of “memes.” A meme is a “cultural replicator parallel to [a] gene,” or, put another way, a “parasite . . . [that] use[s] human brains . . . as

¹⁸⁷ The language of groups “competing” or genes “striving” is obviously anthropomorphic. I mean only that they are subject to forces of natural selection.

[its] temporary homes and jump[s] from brain to brain to reproduce” (Dennett 2003:175, cf. Dennett 1995 and Dawkins 1976). The basic idea behind memes involves applying the general logical structure of Darwinian natural selection beyond the specific context of genes or other physical-biological entities. Genes are relatively complex molecules capable of mutations that can either enhance or diminish the capacity of the gene to replicate in a particular environment. Self-enhancing mutations produce more gene-copies and the mutated genes spread and persist, while self-diminishing mutations eventually perish. Similarly, memes are relatively complex units of culture; “made of information,” memes can be “carried” as contents of mental states or written in a book or stored on a computer or posted on a billboard.¹⁸⁸ Like genes, memes are capable of mutations that can either enhance or diminish the capacity of the meme to replicate in a particular environment. Self-enhancing mutations produce more meme-copies and the mutated memes spread and persist, while self-diminishing mutations eventually perish. Memes can include items as diverse as melodies that get stuck in one’s head, corporate logos, mathematical theorems, cooperative strategies, religious doctrines, habits, biases, artistic techniques, and so on. Any possible “unit of culture” is capable of mutation and subject to forces of natural selection. The most successful memes survive.

Just as genes did not exist on earth until a couple billion years ago, sophisticated memes did not exist until about 50,000 years ago, when certain groups of animals developed brains sufficiently advanced to develop cultural mechanisms for the transmission of information. Daniel Dennett has referred to a “euprimatic revolution” (Dennett 2003: 179), when a new form of primate emerged – a euprimate or “superprimate,” a “hominid with an infected brain, host to millions of cultural symbionts.” The “chief enablers” of this revolution “are the symbiont systems known as languages” (Dennett 2003: 173). Many animals, of course, have primitive forms of culture, and the culture study of animals has become an increasingly important topic within contemporary biology. Birds pass on songs, non-genetically, from parents to their young; and gorillas pass on strategies for hunting and tool use. But the scale of human culture is unique among animals on earth. At first, the development of linguistic

¹⁸⁸ Strictly speaking, one could consider genes to be a subset of the general category meme, since genes, too, carry information. If one pushed the point sufficiently far, one could consider rays of light to be memes, since they carry information about objects that have emitted them. For the rest of this discussion, however, I reserve the term “meme” in general for bundles of information that could be considered “units of [human] culture.” Thus a billboard could be the medium for a meme, insofar as it presents a unit of human culture as a possible content of a human mental state. A gene is not a meme, since, while we can think about it, it is not a “unit of culture.” (The idea of a “gene,” however, is a meme, as is the idea of a meme. □)

capacity served humans' selfish genes. Humans with brains that could host more memes created communities with more advanced possibilities of cultural transmission that were better able to navigate the world in which they lived. Such communities grew and thrived, while communities with less cultural potential died off. Human brains' abilities to generate, host, and transmit memes grew.

But once human brains became efficient meme-creators, mutators, and replicators, memes took on a life of their own. Like parasites, some memes enhance the fitness of their hosts (e.g., hygiene techniques), while others do not (birth control techniques). In some cases, memes that inhibit their hosts' fitness thereby destroy their potential for replication (Shakers' commitment to universal celibacy). In other cases, memes can thrive and replicate even when they do not serve the interests of the genes of their hosts (birth control, again). And some memes that might enhance their hosts' fitness nonetheless aren't very good at replicating (information about foods' caloric content). Of course, memetic fitness and genetic fitness are not wholly unrelated. Memes are only possible because genes that code for meme-friendly brains were more successful than genes that code for meme-resistant brains. Memes sufficiently destructive to their hosts (suicide-for-fun) rarely survive. And memes sufficiently destructive to human genes in general (knowledge of nuclear weapons) could bring the whole memetic enterprise on earth to its end. Moreover, the particular structure of the brain evolved through selection of genes and provides the context for which particular memes will thrive. But these forms of dependence are loose. In general, memes have lives of their own, using human brains as hosts, but promoting the interests of human genes (and even human beings) only as far as that is necessary for their own replicative success.

This relative independence of memes and genes provides a Darwinian way of explaining those aspects of our lives that can seem mysterious from a narrowly gene-centered point of view: art, religion, poetry, and even the sciences are all memes or systems of memes.¹⁸⁹ Creativity in these fields is the result of the tendency of memes, in the medium of the human brain, to mutate. Moreover, because memes often include standards for the adoption of future memes, "successful" memes

¹⁸⁹ Some aspects of human culture – filial piety, for instance – might be so basic and so universal that a genetic explanation seems more plausible than a memetic one. Even basic and universal cultural practices could be the results of either early memetic inheritances shared in common due to common (or interacting) cultural ancestors or convergence upon useful memes by disparate communities. Here there is a close parallel with phenotypic similarities amongst diverse populations of organisms. Often, these similarities reflect common ancestry. In other cases (such as the development of "fins" in both whales and fish) they reflect the convergence of genetically dissimilar organisms towards similar adaptive characteristics.

will be those that conform to the standards of the memetic landscape in which they emerge. The general model of memes as structures that mutate and compete for replication and persistence in human brains can make sense of progress in science; “great works” of art, music and poetry; the delicate balance of tradition and development that characterizes religious traditions; and the tendency of most complex memetic structures to incorporate techniques of education/indoctrination and persuasion/proselytizing. Recognizing the role of memes even helps make sense of how we can find insatiable human thirsts for knowledge and art “for their own sakes,” since these memes are not dependent upon their connections with any other form of success but solely focused on exploiting the distinctive characteristics of the human mind to replicate themselves.

The combination of more sophisticated thinking about evolutionary theory (as in the case of altruism) combined with the addition of memes to the evolutionary framework has given rise to Darwinian-naturalistic accounts of human freedom and morality. Not all Darwinian naturalists think that freedom is something worth saving. Explaining human beings in terms of evolution by natural selection, especially with the addition of selfish genes, makes many think that freedom, and even morality, is simply a relic of scientific ignorance (e.g. Pereboom 2001). But the most sophisticated philosophical appropriations of Darwinism have sought to make sense of freedom and morals. Memes provide a first, crucial tool in freeing human beings from genetically programmed behaviors. Just as Kant emphasized the importance of the “higher faculty of desire” – motivation to act on principles to which we are committed rather than mere instincts – Darwinian naturalists who appeal to memes distinguish humans from other animals based on the fact that we often act in the interest of memes rather than genes. And meme-motivated action has a very different character than gene-motivated behavior.

[A]ccess to memes [has] the effect of opening up a world of imagination to human beings that would otherwise be closed off. The salmon swimming upstream to spawn may be wily in a hundred ways, but she cannot even contemplate the prospect of abandoning her reproductive project and deciding instead to live out her days studying coastal geography. The creation of a panoply of new standpoints is, to my mind, the most striking product of the euprimatic revolution. Whereas all other living things are designed by evolution to evaluate all options relative to the summum bonum of

reproductive success, we can trade that quest for any of a thousand others... (Dennett 2003: 179).

Already, this is a huge step towards both freedom and morality. Human action takes place in the light of memes, reasons that motivate us insofar as we think about them. And moral systems, as complex memetic structures that develop in the context of our natural tendencies toward altruism, can present themselves as standpoints that inform our actions. But even the addition of the memetic point of view does not yet achieve the sort of freedom that we might want, a freedom captured well by Richard Dawkins in the conclusion to his groundbreaking book, *The Selfish Gene*:

We have the power to defy the selfish genes of our birth, and, if necessary, the selfish memes of our indoctrination . . . We are built as gene machines and cultured as meme machines, but we have the power to rebel against our creators. We, alone on earth, can rebel against the tyranny of the selfish replicators. (Dawkins 1976: 215)

So far, we have shown only that selfish genes and selfish memes can pursue their “own interests” independently of each other. Humans need not serve our genes, since we can also serve our memes. But how can we rebel against both genes and memes? And what makes human beings the only creatures that can do this?

The answer to these questions, oddly enough, is a 21st century naturalist version of a key Kantian empirical claim about what distinguishes human beings from animals. For Kant, what “raises him infinitely above all other living beings on earth” is “the fact that the human being can have the ‘I’ in his representations . . . Because of this he is a person, and by virtue of the unity of consciousness through all changes that happen to him, one and the same person” (7:127). In the context of Darwinian naturalism, this “I” is itself a meme, one that has proven particularly adept at self-replication and that opens up a whole new vista of self-understanding and self-control. There are many possible routes for the formation, development, and persistence of the I-meme, more than can be discussed here. One route – suggested by Kant – comes from the conditions necessary for the formation of general concepts. For Kant, the unification of different representations under a general concept requires that one see those representations as belonging to a single “I” (see B131ff.). Thus the complex shift in our mental machinery that makes it possible to move from mere representations of objects to general concepts and thereby to the explosion of memes in human culture is built on a capacity to take oneself as the subject of

one's representations. Another route lies in attempts to coordinate information. Insofar as one's language is limited to claims about objects, it can be difficult to discriminate the perspectives of different knowers. "There is no game in the forest" contradicts "there is game in the forest," but my claim "I did not find game in the forest" need not contradict your claim that you did find game there. A sense of self provides a way to coordinate information. Yet a third basis for the I-meme comes from information-coordination about people, the importance of which is clear in the context of iterated Prisoner's Dilemmas. Humans need ways of communicating about the reliability of other human beings, not merely of other human genes or memes. Humans benefit from evaluating others as persons with fixed characters and holding them responsible for their actions. And as human beings become more sophisticated reasoners, we become capable of deception, which makes the problem of identifying persons (including ourselves) more acute. We develop a sense of self-image, of thinking about how "I" look to others. Cultivating the right image of myself becomes an important social task, and one responds to being held responsible by learning to hold oneself responsible as an efficient way of regulating one's own behavior before others need to step in.

Once human beings have a sense of self, there is no reason that this sense of self must remain motivationally inert. Recall that there were good reasons, from the gene-centered point of view, for the development of memes, but once memes came into the world, they took on a life of their own and could evolve in ways that were not conducive to the fitness of any particular genes. In the same way, the emergence of an I-meme in human brains gives rise to a new kind of entity, a "self" capable of thinking about itself. And this new entity need not serve the interests of the memes that gave rise to it. Moreover, this new entity is precisely a being that sets its own ends. As an entity with a sense of self, this new entity will be capable of higher order desires, reflection on its identity, and even governance of itself by norms – including moral norms – that it "autonomously" endorses. For Darwinian naturalists, this is a sufficient basis for freedom, at least in every sense "worth wanting" (Dennett 1984).

A sufficiently rich Darwinian naturalism thus provides a much better answer to Kant's question than one might expect. Human beings are animals, but we are not "mere" animals. We are social animals that care about one another, expressing sympathy and compassion for those in need and resentment towards those who harm us. We evolved genetic codes that enable the development of a complicated cognitive architecture that makes us hosts to countless "memes." These units of culture mutate and propagate in ways that provide us with a wide

diversity of thoughts, opinions, and practices, artistic creations, religions, and even sciences themselves. Among these memes are moral rules, social and cultural norms, and even that sense of self by virtue of which we regulate our thought and behavior in accordance with what we take to be most important to us. By virtue of our evolutionary history, we have, as Dawkins put it, “the power to defy the selfish genes of our birth.” We are animals but also agents, expressions of genes but also self-expressions, homo sapiens but also wise beings-like-us.

In assessing the possible relationship between Kant and contemporary Darwinism, the most obvious starting point is Kant’s philosophy of biology, where Kant seems to be in trouble. One of the central claims of Kant’s (philosophy of) biology is that

It is quite certain that we can never adequately come to know the organized beings and the internal possibility in accordance with merely mechanical principles of nature, let alone explain them; and indeed this is so certain that we can boldly say that it would be absurd for humans even to make such an attempt or to hope that there may yet arise a Newton who could make comprehensible even the generation of a blade of grass according to natural laws that no intention had ordered. (5:400)

150 years after Darwin laid out a detailed explanation of the origin of species, and 50 years after the birth of the molecular biology that describes precisely how genetic material develops into living things, Kant’s despair about a Newton of a blade of grass seems exaggerated. As Ernst Mayr has put it, “Darwin . . . solved Kant’s great puzzle” (Mayr 1988:58). Today, the assumption of an intentional order in nature is not only unnecessary, but hinders biological progress. Moreover, Kant’s pessimism about mechanical explanations in biology was linked with his use of the concept of “predispositions,” which play crucial roles in his empirical anthropology and are simply taken for granted. Kant offers no explanations of how these predispositions arose and seems to think that the “origin” of the human species is simply irrelevant to anthropology (see 8:110). But Darwinism is precisely the attempt to explain the “origin of species,” and the question of how homo sapiens evolved is central to Darwinian answers to the question, “What is the human being?”¹⁹⁰ At the

¹⁹⁰ For the sake of space, I ignore another Darwinian objection to the question “What is the human being?” Arguably, Darwin’s *Origin of Species* undermines the whole notion of a “species” or “natural kind.” If variations in any particular population of organisms in nature could potentially be bases for what would come to be seen as species-level distinctions, then it seems arbitrary to try to define who “we” human

very least, Kant's indifference to (and even skepticism about) scientific explanations of origins has been shown to be misguided. Contemporary evolutionary biology undermines Kant's pessimism about non-teleological explanation and his specific appeal to innate and inexplicable natural predispositions to explain human (and other living) beings.

In other respects, however, Kant's philosophy of biology is consistent with, and in some ways presciently anticipatory of, present-day Darwinism. For one thing, Kant's appeal to predispositions was a specific and innovative response to the biological debates of his day. In those debates, the main protagonists argued either that biology was reducible to physics, so living things required no special laws in order to be explained, or that all living things were "preformed" in their earliest ancestors (who were created by God). Kant aimed to find a middle ground between these views, arguing that living things are not literally preformed but that explaining them also requires principles that go beyond mere mechanism. Darwinism clearly fits this general model.¹⁹¹ Evolution by natural selection, though not "teleological" in Kant's sense, is a principle for explanation distinct from the mechanical explanations that dominate physics.¹⁹² Even if "natural selection" is in principle explicable in terms of physical forces, evolutionary biologists explain the presence and development of biological features not in terms of physical processes but in terms of the adaptive advantages of these features. Kant was wrong about the specific principles that regulate the practice of biology, but correct that some heuristic principle of a broadly purposive nature is needed in biology. And even if the origins of (human) predispositions are in principle explicable in terms of evolution, Kant was correct that any

beings are. However, given that Darwin himself calls the "immense" difference in mental power "between the highest ape and the lowest savage" (Darwin 1981: 34), for the present chapter I will take for granted that, for practical purposes, we can make a distinction between humans and other organisms, even if the theoretical concept of a fixed "human species" is problematic.

¹⁹¹ Two further features of Kant's biology connect it even more closely with contemporary evolutionary biology. First, Kant emphasizes that the need for teleological explanation is "regulative" and not "constitutive," which means, for him, that despite his claim that humans never will discover non-teleological explanations of living beings, we are "summoned . . . by reason" to "the greatest possible effort, indeed boldness, in attempting to explain [living beings] mechanically" (5: 429). Second, the details of Kant's account reflect a sophisticated 18th century attempt to articulate a biological methodology that would take into account both the apparent heritability of traits and the ability for new (heritable) traits to emerge. Kant noted that living things are susceptible to physical forces, but that they seem to develop according to heritable internal principles. The best explanation for this, Kant suggested, is that the environment affects the expression of predispositions that are passed on from parents to their offspring. Kant even suggested that these patterns of expression can become hereditary, such that predispositions can become inert or substantially modified over what we would now call a process of "evolution."

¹⁹² The issue of whether teleological explanation itself is still necessary for biology remains a live issue, so Kant may have been even closer to the mark that this section suggests. (For contemporary discussions of teleology in biology, see Allen et. al. 1998 and Ariew et. al. 2002.)

explanation of biological characteristics depends upon seeing how environmental conditions affect the expression of inherited (and not immediately explicable) “predispositions” for those characteristics.

But what of the further claim, that an understanding of the evolution of human beings (and their predispositions) provides the basis for a sufficient answer to the question, “What is the human being?” Did Kant, by ignoring the question of origins, pass over the best and most adequate answer to the question that sums up the whole of philosophy?

I think not. From a Kantian perspective, the question “What is the human being?” can be answered transcendently, empirically, and pragmatically. Kantians must acknowledge that evolutionary biology greatly enriches Kant’s empirical anthropology by showing the origin of humans’ natural predispositions, and the accounts of these origins can even help explain the nature of those predispositions. Because empirical anthropology provides empirical insights needed for pragmatic anthropology, revisions and additions to our empirical picture of human beings can enrich or modify pragmatic anthropology. In addition to the many pragmatic, medical uses to which genetics, for instance, has been put, recognizing what makes memes thrive in particular contexts can help us influence others and avoid unwanted manipulation of ourselves. But for Kant, *transcendental* anthropology provides the most fundamental answers to the question, “What is the human being?”, answers that not only get to the root of who we are but that provide the norms that orient pragmatic anthropology. Thus the contest between Darwinian naturalism and Kant must be decided around two core issues:

- (1) Does evolutionary biology provide good grounds for challenging either Kant’s threefold division of anthropology or his prioritization of transcendental over empirical anthropology?
- (2) Does evolutionary biology provide a more adequate approach than Kant to transcendental anthropology, and/or does it provide good reasons to challenge Kant’s transcendental anthropology?

Related to these two questions is a third, one that is central both to Kant’s anthropology as a whole and to recent philosophical justifications of evolutionary naturalism:

- (3) Is the notion of human freedom allowed within evolutionary biology sufficient?

In the rest of this section, I argue that the answer to the first two questions is no. Evolutionary biology does not give good reasons to

conflate transcendental with empirical anthropology, and it fails to provide an adequate transcendental anthropology of its own. I reserve my discussion of (3) until section IV of this chapter.

First, what is the relationship between transcendental and empirical anthropology? It can often seem as though evolutionary biologists deny the need for and possibility of transcendental anthropology altogether. In explaining his naturalism, for example, Daniel Dennett insists that the role of philosophy today is little more than systematizing the insights of empirical sciences. Here the notion of an a priori human science can just seem absurd. But Dennett does recognize that there are distinct perspectives that one can take on human beings (and other things). Even with respect to primitive forms of “life” in a computer simulation, Dennett insists that “our simplest doers have been reconceptualized as *rational agents* or *intentional systems*” such that “we can move back and forth between the . . . God perspective [from which intentionality is the product of other forces] and the ‘perspective’ of . . . God’s creations [in which intentionality is basic]” (Dennett 2003:45). The “meme-centered” point of view and especially the self’s point of view can be explained naturalistically but are not themselves “naturalistic” perspectives. And Dawkins, Dennett, and others rightly insist that once reflection and self-image enter the scene, human beings are capable of asking for reasons and reflecting normatively on what to think and do.

Some memes surely enhance our fitness, making us more likely to have lots of descendants (e.g. methods of hygiene, child-rearing, food preparation); others are neutral – but *may be good for us in other, more important regards* (e.g. literacy, music, and art)—and some memes are surely deleterious to our genetic fitness, but even they may be *good for us in other ways that matter more* to us (the techniques of birth control...). (Dennett 2003: 177, emphasis added).

Despite initial appearances to the contrary, there is no dispute between Kant and the most prominent evolutionary naturalists about whether there is a normative perspective “from-within.” Kant and Dennett both acknowledge that human beings can be studied as empirical objects in nature, and both recognize that the laws that explain the development of human selves are not identical to the rules that govern those selves from-within.

Nonetheless, Kant and Dennett fundamentally differ about the relative priority of transcendental and empirical anthropology. While Dennett explains the legitimacy of transcendental anthropology on the

basis of an empirical account of its evolution, Kant explains the legitimacy of empirical anthropology on the basis of a transcendental account of its justificatory basis. Thus Dennett sees “the creation of a panoply of new standpoints” as “the most striking product of the . . . [biological] revolution” that gave rise to human organisms (Dennett 2003: 179), while Kant would see the ability to give an evolutionary account of human cognition as one of the most striking results of applying our causal way of thinking about the world to the case of ourselves. Dennett sees the biological standpoint as, fundamentally, the true standpoint, the “God perspective” (Dennett 2003: 45). Kant sees evolutionary biology as a standpoint of empirical cognition, which gets at one kind of truth, the truth about the world-as-we-experience-it. Fundamentally, then, the difference between Kant and Dennett relates to the status of science as such.

Philosophical naturalists like Dennett tend to be strong scientific realists. A strong scientific realist is someone who takes natural science, at least ideally, to describe the truth, the whole truth, and nothing but the truth. A strong scientific realist need not think that the current state of science gets everything correct, but insofar as our science fails to tell the truth, the whole truth, and nothing but the truth, it needs to be improved. By contrast, Kant is a sort of limited scientific realist, in that he takes an ideal science to lay out the truth and nothing but the truth, but for Kant, science specifies the truth only about the world-as-we-can-experience-it. Scientific claims are claims that human beings, given our structure of cognition, should believe about the world. This has two important implications for Kant’s appropriation of evolutionary biology (or any science). First, even though evolutionary biology is an empirical science, it is possible only because of certain a priori structures of human cognition, and these can be studied independent of particular empirical results. If (counter-factually) evolutionary biology were to find it impossible to explain the evolution of causal reasoning, for example, this would not undermine the legitimacy of causal reasoning, since causal reasoning is an a priori condition of the possibility of any empirical science at all. Second, it means that evolutionary biology is itself subject to transcendental critique. If (again, counter-factually) Kant’s transcendental philosophy gave some reason to call into question the methodology of evolutionary biology, we would have to give it up as a legitimate way of gleaning knowledge about human beings, regardless of how otherwise handy it seems to be.¹⁹³

¹⁹³ For an example of this from Kant’s own work, see A173-4/B215-6 (quoted in chapter nine, p. xxx).

Should we be strong scientific realists? There are at least two reasons for skepticism here.¹⁹⁴ The first relates to Kant's central argument for his Copernican turn. Science operates in the context of assumptions that guide inquiry and restrict the scope of scientific explanation. To some extent, these assumptions are justified in retrospect, by their success. Hypotheses are "confirmed" when they bear fruit in terms of predictive or explanatory success. Ultimately, though, even the claim that predictive success is an indicator of truth is a mere assumption. And science operates with numerous heuristics ("look for adaptive advantages of distinctive features" and "as much as possible, explain similar effects by appeal to similar causes") and restrictions ("do not appeal to divine explanations" and "the future cannot cause changes in the past") that are not empirically tested but nonetheless constrain scientific explanation. As Kant argues, some of these scientific assumptions are simply impossible for human beings to question. We explain changes in terms of causes, for example, and we assume that nature is uniform. We might add that insofar as we engage in scientific explanation of the world, we must treat hypotheses that have a high degree of predictive success as more likely to be true than those – such as divine intervention – that have no predictive value at all. But all of these standards are rooted in the (transcendental) nature of human cognition; they all reflect the necessary conditions for humans to understand an empirical world. (A god probably would not need to appeal to predictive success to confirm hypotheses.)

Evolutionary biology is a human science. As far as we can tell, neither gods nor animals think about the world in terms of evolution by natural selection, and there are very basic assumptions underlying evolutionary biology that cannot be seriously questioned. What are we to make of the importance of substantive and methodological assumptions in human science? We could take these assumptions as self-evident, but doing so imports a dogmatic rationalism into science. We could take them as purely arbitrary, but this would undermine any justification for scientific realism. We could simply not worry about them; who doubts, after all, that a theory with massive predictive success is at least closer to the truth than one that consistently fails to make accurate predictions? This approach makes it psychologically possible to sustain a commitment to strong scientific realism, but provides no justification for that realism. Kant provides a better approach. Given that certain conditions seem to be necessary in order for humans to experience the world, we can simply take these conditions to be true of the world we

¹⁹⁴ In the next chapter, I emphasize a third reason in the context of recent historicist accounts of science, and in chapter ten, I look at existentialist critiques of scientific naturalism.

experience. This is a pretty strong sort of realism, in that it takes the best scientific theories of the world to be true, but it is limited in that it admits that the world of which these theories are true is the world as we experience it. While preserving a substantial commitment to scientific knowledge of the world, Kantian scientific realism rejects the God's-eye point of view assumed by strong scientific realists such as Dennett.

In itself, this Kantian scientific realism is *consistent* with strong scientific realism (what Kant would call "*transcendental* realism") in that one could simply take on faith that the world we experience exhausts all that there is. Just as Kant's immediate successors argued for the "neglected alternative,"¹⁹⁵ contemporary naturalists might affirm that the basic presuppositions of empirical knowledge are both necessary structures of human cognition and sufficient for exhausting all there is to the world. Moreover, one might wonder what the point of limiting one's scientific realism could be. Even if there is some mysterious world-beyond, if we can neither experience that world nor have any way of getting knowledge about it, what difference does it make? Even if science does not exhaust "the whole truth," if it exhausts everything true that humans can know, then shouldn't we – humans that we are – just go ahead and be strong scientific realists?

Perhaps. But Kant gives reasons for rejecting this "neglected alternative," of which the most important here is that scientific explanation is merely one perspective that human beings take on the world. Scientific descriptions and causal explanations of the empirical world are constrained by certain basic concepts and methodological assumptions. But humans must also make sense of the world from within the standpoints of practical deliberation about actions and epistemic deliberation about scientific theories. For evolutionary biology to provide a sufficient answer to the question "What is the human being?" it must make sense of these standpoints. That is, strong scientific realism – the view that science provides the whole truth – depends upon evolutionary biology providing an adequate transcendental anthropology. And this is something that evolutionary biology fails to do.

Evolution is very good at explaining how various human predispositions evolved. But these stories fail to reveal the transcendental structure of our faculties of cognition, feeling, and volition. For example, we can tell stories about why humans have cognitive structures that make us think that $2+2=4$, but this neither shows whether the thought that $2+2=4$ is actually justified nor reveals

¹⁹⁵ See chapter six, pp. xxx.

the conditions of possibility of such justification. Neither genetic nor memetic success can explain why $2+2=4$, even if they can explain why we *believe* that $2+2=4$. Similarly, we can tell stories about how kin selection and cognitive evolution gives rise to genetic characteristics that contribute to a propensity to endorse certain ethics-memes, but this cannot show whether we are right to endorse those memes, nor what the transcendental conditions of possibility of choice really are (that is, what is implied in our taking ourselves to be responsible for something).

And when it comes to these sorts of normative claims, evolutionary approaches are notoriously unhelpful. When Dennett notes that memes that are deleterious from the point of view of survival and reproduction “may be good for us in other, more important regards,” he says nothing about *why* those other regards are or even may be more important. And when he asks “whether or not morality itself is a feature we should try to preserve in our societies” (Dennett 2003: 279), there is nothing in his evolutionary account that can answer this question. An evolutionary account might be able to explain why birth control *does in fact* matter more to us than genetic fitness, but Dennett does not even attempt to show how it can explain what makes certain goods (he offers examples like literacy and music) genuinely more important than reproductive success. The reason for this failure is one that Kant rightly emphasizes. From-within, when one is actually trying to figure out what to believe, feel, or do, one looks not for explanatory causes that could predict future beliefs or choices but for justificatory reasons of them. Once there are beings in the world who are capable of reason-guided reflection, those beings take standpoints on the world from within which causal explanations are insufficient. If one is trying to decide whether to read an edifying book or help a friend move into a new apartment, reading a psychological assessment of oneself that explains that one will avoid reading books whenever one has something else to do will not actually help one make one’s decision. And that is because, in itself, the psychological assessment is not a reason. It might, of course, give one a reason; one might rebel against the report just to assert one’s independence, or one might use the report as part of a justification for reading the book, since it is “inevitable anyway.” (Kant actually thinks that part of the appeal of determinism is that it allows us to abdicate responsibility and thereby act on desires and against what is really normatively required.) But one needs some basis for taking this psychological assessment as a reason for rebellion or complacency, and the report itself cannot provide this basis.

The problem of justification is not limited to choices about actions; it arises even for the practice of science itself. Evolutionary biology fails

to provide a justification for the very cognitive practices that it employs. Alvin Plantinga has emphasized this point against Dennett's evolutionary naturalism:

Darwin's dangerous idea is really two ideas put together: philosophical naturalism together with the claim that our cognitive faculties have originated by way of natural selection working on some form of genetic variation. According to this idea, then, the purpose or function of those faculties (if they have one) is to enable or promote survival, or survival and reproduction, more exactly, the maximization of fitness (the probability of survival and reproduction) . . . [T]he probability that our cognitive faculties are reliable (i.e., furnish us with a preponderance of true beliefs) on Darwin's dangerous idea is either low or inscrutable . . . If so, then it also gives [one] a reason for doubting any beliefs produced by those faculties. This includes, of course, the beliefs involved in science itself. (Plantinga 1996)

Insofar as memes allow human beings to transcend "selfish genes," it is not fair to see beliefs as serving merely to promote survival or reproduction of genes, but Plantinga is correct that nothing about evolution itself provides a reason to believe that beliefs that arise through processes of evolution are justified or true.¹⁹⁶ Kant helps us see (contra Plantinga) that evolutionary explanation is compatible with an epistemology that could justify knowledge-claims, but this does not show that evolution is sufficient for such an epistemology. The empirical account of human beings provided by evolutionary theory must be supplemented by something like Kant's transcendental anthropology, and because this transcendental anthropology provides the conditions of justification of science itself, it must be viewed as more fundamental than evolutionary biology.¹⁹⁷ Moreover, regardless of how sophisticated a description of human beings – even as empirically knowable "selves" or "agents" – we have, this description is insufficient to justify adopting any

¹⁹⁶ Moreover, evolutionary theory may have at its disposal accounts of human cognition that could provide justification for scientific realism. See, e.g., Popper 1972: 261ff. and Quine 1969: 126, but cf. Plantinga 1993.

¹⁹⁷ I'm not here denying the possibility of a sort of Quinean holism or reflective equilibrium model (see Quine 1951). Even in those models, there are forms of logical priority, but what is logically secondary can lead to revisions in what is prior. Given the structure of the relationship between the transcendental conditions of possibility of empirical science and the specific content of biology, any rejection of the former would involve a rejection of the latter. Now discoveries within empirical sciences might, for Quine, provide reasons for rejecting our transcendental conditions of science and the sciences based on them, but the result of this wholesale revision of our epistemic landscape would not involve a subordination of transcendental philosophy to evolutionary biology but a rejection of both in favor of some new way of thinking about the world.

particular beliefs or choices without some reason to make use of this description in some particular way. Because evolutionary biology fails to provide for a transcendental anthropology that is nonetheless necessary, strong scientific realism is false. Evolutionary biology might tell the truth and nothing but the truth, but it does not tell the whole truth.

In itself, evolutionary naturalism fails to provide a transcendental anthropology. Does it raise fundamental problems for Kant's attempted transcendental anthropology? Again, I think not. As in the case of contemporary neuroscience, Kant's distinction between empirical and transcendental anthropology largely insulates his transcendental anthropology from empirical objections. For example, Dennett has criticized Kant's moral theory in the context of his account of the evolutionary origin of moral judgments:

Kant held that [pure, emotionless] judgments are not only the best sort of moral judgments, they are the only sort of judgments that count as moral judgments at all. Enlivening reflection with base appeals to emotion may be fine for training children, but the presence of those training wheels actually disqualifies their judgments for moral consideration. Is this perhaps a case in which holding out for perfection – a job-related disability in philosophers – conceals the best path? (Dennett 2003: 213)

The question-mark here is apt, since Dennett actually provides no reason at all for thinking that Kant is misguided to “hold out for perfection.” How would one answer this question? Surely not by appealing to our evolutionary history, nor even to the role that emotions actually play in most (even all) judgments that are considered “moral.” Rather, one must look at the structure of volition from-within, asking what would justify “holding out for perfection” or looking for a better path. And in fact, Kant does enter into this sort of reflection. In his *Groundwork*, he argues that while human beings do in fact act for the most part from what we might call “emotions,” when reflecting upon what we ought to do, we do not actually think that a commitment to do what one feels like doing is morally praiseworthy, even if one feels like doing the right thing. Kant might be wrong about this, but if he is, it is because he misread what volition looks like from-within, not because he failed to trace the evolutionary origin of moral judgment.¹⁹⁸

¹⁹⁸ Kant also provides good from-within grounds for a stance towards moral life that both holds out for perfection and pursues a better path. In response to radical evil, Kant insists that one must not give up moral perfection as an ideal, but also that one can justify moral hope on the grounds of progress towards this ideal rather than perfect conformity with it.

As in the case of neuroscience, contemporary evolutionary biology fleshes out empirical anthropology far beyond Kant's expectations. This not only enriches our empirical self-conception but provides valuable insights that can be used in pragmatic anthropology. But Kant's transcendental anthropology – or something like it – is necessary to explain the science's conditions of possibility and to guide how which empirical knowledge is put to use to justify norms in terms of which humans ought to think, feel, and act.

III. Contemporary Psychology

The previous sections drew attention to developments in biology with significant impacts for understanding ourselves as human beings. Many recent advances in psychology are rooted in these biological developments. Psychologists today make extensive use of neuroscience and evolutionary modeling in studying the human mind. For example, one can make claims about the mental processes involved in various activities through scanning the brain and noting which areas are most active while subjects are engaged in various tasks, and psychologists have used studies of animals– especially those most closely related to us – to gain insight into the way human brains work. But psychology has also made significant progress as a science distinct from biology, most notably through increasingly sophisticated experimental methodologies that provide evidence for claims about humans' mental lives. Because Kant developed a detailed empirical psychology, it is natural to compare Kant's psychology with contemporary psychological methods and theories. Moreover, as in the case of neuroscience and evolutionary biology, contemporary psychology has been a source for naturalist approaches to human beings. And recently, philosophers have appealed to specific findings in psychology that seem to raise problems for Kant's anthropology. This section starts with a brief Kantian discussion of methods and models within contemporary psychology and then turns to two ways in which contemporary psychology has fed naturalist philosophical accounts of human beings in ways that seem to threaten Kant's epistemology and moral theory.

Contemporary psychology has made considerable strides towards reappropriating a broadly Kantian approach to the mind. In the mid-20th century, the dominant approach to psychological research was behaviorism. Promoted especially by B.F. Skinner, behaviorist

psychologists assumed that human mental life was reducible to externally observable behaviors. In its most extreme form, the human mind was seen as a mere stimulus-response machine, and psychology was the science of classifying stimuli and responses. During the past 50 years, however, psychologists have regained an interest in the mind as such. Partly this has been for experimental reasons; one famous study on rats suggest the reality of “latent learning,” where animals make evident that they knew things that were not normally expressed in behavior (see, e.g. Tolman and Hoznik 1930 and Rescorla 1991). Partly, though, the shift away from behaviorism comes from a more obvious source. Human behavior is not a brute fact, nor should scientists be limited in explaining it to laying out a sequence of physical states.¹⁹⁹ Certain very simple human responses might be explicable in terms of innate or conditioned responses to stimuli. Humans may innately respond to the perception of a yawn with another yawn, and can be taught to yawn on cue if sufficiently conditioned. But even to explain something as simple as why one runs out of a flaming building (see Pinker 1997:62) or why one looks for one’s own car in a parking lot, one must appeal to beliefs and desires. And more complicated aspects of being human seem utterly inexplicable without appealing to mental states as such. Imagine trying to distinguish, based purely on conditioned responses, between a person who marries for money, another who marries because she doesn’t want to die old and alone, another who remains unmarried to avoid a messy divorce later, and a last who doesn’t marry in order to have a good career. Explanations in terms of mental states are straightforward and predictively successful; those in terms of external inputs and behavioral outputs alone are hopelessly insufficient. Thus the study of mental states as such has become a mainstay of psychology (again).

For Kant, this is welcome. Kant’s psychology is a systematic study of mental states and the relationships between them. Like contemporary psychologists, Kant is willing to explain some behaviors in terms of conditioned responses, but like most psychologists today, he appeals to more complicated mental states to explain most human behavior. Moreover, like contemporary psychologists, Kant is not content with casual folk psychological explanations of behavior. Although he resists behaviorism, Kant seeks law-like relationships among clearly delineated types of mental states. And this brings out a further important parallel between Kant’s psychology and contemporary psychology. Increasingly,

¹⁹⁹ Even the most ardent behaviorists (such as Skinner himself) implicitly made use of descriptions of stimuli and responses that appealed to internal psychological states. For discussion, see Dennett 1978, Fodor 1968, and Pinker 1997.

psychologists today reject “blank-slate” approaches to human mental life in favor of an approaches that look more like Kant’s taxonomy of basic powers.²⁰⁰ As one philosopher of psychology has put it, “one of the major insights of [contemporary psychology] has been the extent to which we depend upon a natural cognitive endowment, which assigns processing tasks to modular structures with quite specific and restricted domains and inputs” (Botterill and Carruthers 1999:50; see too Fodor 1983, Pinker 2002). This “modular” approach to the mind rejects the reduction of all mental processes to a few simple potentials that develop in different ways through human learning. Just as Kant divided the mental into irreducible but interacting “powers” and “faculties” that operate by different causal laws, contemporary psychologists study the human mind as a set of interacting “modules” that perform different tasks in bringing about human thoughts and actions. Moreover, like Kant, psychologists distinguish between the biological bases for mental modules (what Kant would call predispositions) and the fully-formed mental modules (powers) that emerge when these bases are able to develop in particular contexts (Botterill and Carruthers 1999:96).

There are still important differences between Kant’s basic powers and modern mental modules. Kant’s approach was situated within a metaphysical model of substances interacting by means of powers, while the modern approach sees modules not as distinctive active properties of a substance but as mental functions rooted in the evolved architecture of the brain. Kant’s “powers” were also fairly commonsensical, and one distinguished between them using a broadly introspective approach. And Kantian powers are domain-general, in that a given power covers a wide range of possible contents. (Reasoning about baseball statistics and reasoning about the reliability of one’s friends are both rooted in the same basic power: reason.) The result is a cognitive map including powers of vision and hearing, of judgment and reason, of feeling and desire. By contrast, modern modular accounts of the mind assume that most modules are unconscious and can be distinguished by studying developmental evidence (how children’s cognition is able to progress in some ways before/without developing in other ways), psychological damage (where one module’s activity is inhibited but not others’), and even brain-scanning (looking for regions of the brain active in various tasks). And domain-specific modular accounts of the mind are increasingly the norm. Thus the lists of mental modules in contemporary psychology look quite different from Kant’s taxonomy of powers,

²⁰⁰ For the sake of simplicity, I am conflating a blank-slate approach with an anti-modular approach. The most common forms of anti-modularism are black slate approaches, though there are ways of avoiding both approaches to mind, such as connectionism and associationism. For further discussion, see Pinker 1997.

including specific modules for the perception of color, the perception of shape, the detection of rhythm, and the recognition of other people (Fodor 1983: 47-8). While Kant shares with contemporary psychology a commitment to broad heuristics according to which structures of the mind are distinguished based on empirical evidence, the much broader range and types of evidence today has given rise to a taxonomy that is both different from and more fine-grained than Kant's.

Kant could accept most of these modifications of his view. While Kant sought to reduce the number of basic powers to as few as possible, he would accept that mental structures that seemed unified to him – such as vision, say, or reason – involve irreducible sub-components. As some recent philosophers of psychology have noted, “failure to draw a distinction is not at all the same thing as denying that there is a distinction” (Botterill and Carruthers 1999:73). Kant did not draw all the distinctions of contemporary modular psychology, but his empirical methodology lends itself to a willingness to admit sub-distinctions within his overall faculty psychology. And many of Kant's most important distinctions, such as his tri-partite conception of mental faculties or his distinction between higher and lower cognition, have been vindicated by recent psychological research.²⁰¹ The fact that many of these processes are unconscious need not pose intractable problems for Kant's empirical psychology.²⁰² While there is a philosophical challenge in making sense of what it means for a mental state to be unconscious, Kant already made important steps in this direction. Natural propensities need not be conscious, lower faculties are not conscious in a reflective sense, and

²⁰¹ With respect to the tripartite soul, recent research on the chemical bases of pleasure suggests that pleasure is often but not always connected with desire. Among the most interesting work in this regard is psychological research suggesting that there is a sort of pre-cognitive feeling of pleasure and pain that is (sometimes) transformed into something with cognitive and volitional import. The evidence for this distinctive feeling of pleasure comes from at least two sources. From neurobiology, the recognition that certain chemical changes in the brain – most notably, the release of endorphins – is consistent across a wide variety of pleasurable emotions, suggests that there is some basic psychological state shared in common between these emotions. Situationist psychological research confirms this through studies that show that the same physiological states will be interpreted as different emotions depending upon situational cues. For example, in the Capilano Suspension Bridge Experiment, two investigators found that male subjects who have just crossed over a rickety suspension bridge see themselves as more attracted to female interviewers than those who have crossed over a safe and secure bridge; the subjects interpret the high heart rate caused by their anxiety as sexual attraction (see Dutton and Aron 1974).

²⁰² Given that Kant's psychology is based largely on introspection, one might think that Kant cannot allow unconscious mental states to play a role in it, since such states are, by definition, unavailable to introspective awareness. But while introspection enjoys a central place in Kant's psychological method, he also makes inferences from that which is directly available to introspection and posits psychological features of which we are not immediately conscious. There is therefore nothing in Kant's psychological methodology that precludes him from accepting unconscious mental states insofar as these are needed to make sense of what is observed (through introspection and perception of behavior).

Kant was perfectly willing to allow for unconscious physical processes underlying humans' mental states. To make the further move that there could be particular psychological processes that operate like conscious mental states but without consciousness is a step beyond Kant, but not one he would have to reject.

Methodologically, however, two shifts have occurred over the past 100 years that raise questions about Kant's conception of the nature of empirical psychology. First, Kant insisted in his philosophy of science that psychology "can . . . never become . . . a science" (4:471). Given the progress over the past 200 years, one might ask: Is contemporary psychology a genuine science? Second, Kant's own empirical anthropology is rooted in introspection, but introspection is widely regarded with suspicion in contemporary psychology. One might wonder, then, whether contemporary psychology has shown the fruitlessness of the fundamental bases of Kant's empirical psychology. With respect to both issues, Kant is actually much closer to contemporary psychology than it might seem, and where they diverge, Kant raises genuinely important issues that contemporary psychology should address.

With respect to the first issue – the scientific status of psychology – there is virtually no real disagreement between Kant and contemporary psychologists about the nature of psychology. Psychology today, as for Kant, is a wholly empirical study of the human mind that aims to lay out the structure and explain the development of human mental structures. Psychologists aim to explain the full range of human mental processes using the simplest general structures. Specific thoughts and actions as well as the mind's underlying structures should be explained in accordance with causal laws. On all of these points, Kant's description of psychology fits contemporary practice. But Kant calls this sort of discipline a "natural history of the mind" that is only a "science" in a loose sense. For Kant, science strictly speaking must have an a priori foundation, and Kant insists that no such foundation can be found for psychology. As far as contemporary psychologists indicate, Kant is correct.

There are two ways, however, in which contemporary psychology is more "scientific" than Kant supposed possible. First, psychology has made substantial progress towards rooting psychological explanation (especially of the origin of human mental structures) in biological explanation. Kant, too, situated his psychology in the context of biology, but – as we saw in the last section – he underestimated the extent to which a causal account of the origin of human beings could be given. In that sense, evolution-informed psychology goes beyond Kant's

expectations for the science. Second, Kant insisted that “there can be only so much *proper* science as there is *mathematics* therein” (4:470) and that “mathematics is not applicable to” psychology (4:471). But some forms of contemporary psychology are highly mathematical. In addition to mathematical models of brain activity, contemporary experimental psychology is largely dependent upon statistics to describe, organize, and interpret data. Nonetheless, while this mathematization of psychology is important and unexpected by Kant, the kind of mathematics of which psychologists make use is not one that confers scientific status in Kant’s sense. For Kant, mathematics makes physics scientific because it allows the physicist to make *a priori* claims at the foundation of physics. (For example, Kant uses the fact that the surface area of a sphere is proportional to the square of the distance from the center of the sphere to argue – *a priori* – for an inverse-square law for gravitational force.) While mathematical laws of statistics help psychologists process empirical data more effectively, they provide no *a priori* insights into the nature of the mind.

With respect to the second issue – introspection – the divergence between Kant and psychology today might seem more important. While Kant insisted that empirical anthropology “is provided with a content by inner sense” (7: 398, cf. 25:252, 863-5), contemporary psychologists often disparage “introspection” as an outdated and unscientific approach to studying the mind. Moreover, there is empirical evidence questioning introspection as a methodological tool. The most famous article to this effect concludes that “there may be little or no direct introspective access to higher order cognitive processes” (Nisbett and Wilson 1977: 231). The evidence comes from countless studies in which subjects questioned about the causes of their own beliefs or actions fail to accurately report on these causes. In one such study, subjects were invited to evaluate the quality of various consumer products (4 different nightgowns in one iteration of the study, 4 identical nylon stockings in another). The result was “a pronounced left-to-right position effect, such that the rightmost object in the array was heavily over-chosen. For the stockings, the effect was quite large, with the rightmost stockings . . . preferred . . . by a factor of almost four to one.” But although the position of the products was clearly a factor the choices of at least some subjects,

when asked about the reasons for their choices, no subject even mentioned spontaneously the position of the article . . . and when asked directly about a possible effect of the position of the article, virtually all subjects denied it, usually with a worried glance at the interviewer suggesting that they felt either that they had

misunderstood the question or were dealing with a madman. (Nisbett and Wilson 1977: 243-4)

Introspection is so unreliable that even when people seem to be accurate about what is really moving them to make a particular judgment or decision, the authors conclude (and have evidence to back up) that this accuracy is based on an inference from behavior and context to internal states, precisely the same sort of inference that would be made by an external observer.²⁰³

In fact, however, Kant and contemporary psychology are far closer than they seem, even with respect to introspection. For one thing, the move away from behaviorism requires at least some appeal to introspection. Even the claim, for example, that “subjects are . . . unaware of the existence of a stimulus that importantly influenced a response” (Nisbett and Wilson 1977: 231) assumes that the reports (and/or the “worried glances”) of subjects are reliable indicators of the subjects’ “awareness.” But the need for some reliance on introspection goes much deeper insofar as non-behaviorist psychologists take this research to have implications for how mental states themselves can be studied. As Nisbett and Wilson put it,

The explanations that subjects offer for their behavior in insufficient-justification and attribution experiments are so removed from the processes that investigators presume to have occurred as to give grounds for considerable doubt that there is direct access to these processes. (Nisbett and Wilson 1977: 238)

“Insufficient-justification” experiments are those in which subjects are given very small inducement (say, \$1) to perform unpleasant tasks, and subjects typically report that tasks to be more pleasant than those who are given stronger inducement (say, \$20). Investigators typically explain this in terms of a need for subjects to see themselves as having acted reasonably. Since a small inducement is not a good reason for performing a very unpleasant task, subjects convince themselves that the task is not really that unpleasant. But most subjects, even when explained this hypothesis, cannot see themselves as having been moved by this consideration. Introspection fails to pick out relevant mental processes. Importantly, however, the investigators’ attribution of these mental processes to their subjects is based, at least indirectly, on

²⁰³ I really mean *precisely* the same sort of inference. Studies have shown that those participating in the experiments and external observers make the same judgments – either accurate or inaccurate – about participants’ internal states. Apparently, introspection does not introduce any significant new information about what is going on. (See references in Nisbett and Wilson 1977 and Bem 1967.)

introspection. Investigators do not merely think that \$1 has a magical property of reducing unpleasantness while \$20 lacks such a property. Instead, they think about what would “make sense” of the different responses of subjects. But this judgment of “making sense” is based on a very general sort of introspection, one’s long experience with the sorts of considerations that motivate one to think and act in certain ways.²⁰⁴ Even if people are often wrong about the particular motives for particular reactions in particular cases, introspection still seems to be an effective and even necessary tool for discerning what general sorts of mental states there are and how, in general terms, these mental states interact with one another. Insofar as psychologists make claims about mental states as such, rather than seeing people as mere stimulus-response mechanisms, they must include at least some appeal to introspective awareness.²⁰⁵

Of course, the need to appeal to introspection in this general way does not alleviate the very real problems to which these experiments draw attention. But here it is important to recognize that Kant, too, was acutely aware of the limits of introspection. In his *Anthropology*, Kant lists “considerable difficulties” that face any psychology seeking to “trac[e] everything that lies hidden in” the human mind. He specifically mentions both dissembling, where one “does not want to be known as he is” and a sort of embarrassment that make it impossible to show oneself as one really is. And with respect to many mental processes, Kant points

²⁰⁴ Cf. NW 1977: 248, where Nisbett and Wilson rightly point out that many of these theories can come from cultural inheritances. We inculcate cultural expectations about how people act and why, and these “a priori causal theories” affect our judgments about motivation in particular circumstances. But this only provides a partial explanation, since these cultural theories must originate and evolve somehow. Most plausibly, the origin of such theories is introspective; people pay attention to and generalize their own motivational structure, and then refine their theories through ongoing interaction with others, including others’ introspective reports.

²⁰⁵ Kant’s approach to human beings can also help shed light on an important possible problem with the structure of these self-reporting experiments. When asked why one did something, one can interpret this question either as an empirical-causal one or as a transcendental-justificatory one. A question like “Why did you prefer that stocking over the others?” is most naturally seen as addressing a person from-within, asking for a justification rather than a causal explanation of the judgment. As a justification, the fact that the stocking was furthest to the right is, frankly, mad. (Hence the “worried glance.”) Given Kant’s careful distinction between introspection – where one looks from-without at one’s internal states – and the from-within perspective of justification, one might reasonably conclude that many of these psychological experiments prompt, not introspection, but self-justification. One interesting experiment in this regard (Lepper et. al. 1970) even suggests an important connection between the justificatory structure used to “explain” one’s decision and the deliberative context of the decision itself. How one “explains” one’s actions seems to depend upon what one needed to think at the time of those actions to justify them from-within. Exciting introspection (and cultivating a discipline of careful introspection) is something that requires specific instructions that can be missing from these psychological experiments. True introspection – if genuinely elicited – might be more effective than these experiments suggest.

out that “when the incentives are active, he does not observe himself, and when he does observe himself, the incentives are at rest” (7:120-121, 398-9). As with Kant’s empirical anthropology in general, contemporary psychological research has provided substantially more specification of and evidence for the difficulties with introspection to which Kant drew attention. But Kant was not so naïve about introspection that he would find this recent psychological research surprising, nor is psychology today capable of doing without at least the general and constantly corrected introspection that Kant saw as lying at the heart of empirical anthropology.

Kant’s approach to psychology, then, is broadly compatible with contemporary “scientific” psychology. But as in the cases of neuroscience and evolutionary biology, philosophers have appropriated psychology in the service of a thoroughgoing naturalism about human beings. Psychological naturalism need be neither materialist nor reductionist, which alleviates some of the problems directed against neuroscientific naturalism in section one.²⁰⁶ Since multiple realizability is a problem only for naturalisms that attempt to reduce the psychological to the physical, it not a problem at all for psychological naturalism. Qualia and intentionality might pose problems for psychological naturalism, since these properties of mental states seem to manifest themselves primarily from-within cognition, but even mental qualia and intentionality can become objects of introspective awareness and in that sense could be incorporated into a scientific psychology. The biggest problem for psychological naturalism is the problem of normativity, the justificatory status of reasons from-within. To solve this problem requires providing psychologically-rooted, naturalized epistemology and ethics.

Naturalizing epistemology has become a thriving research program among philosophers today (see Feldman 2001). At its most extreme, such a view involves the commitment to wholly replacing epistemology with psychology. As W.V. Quine famously put it,

Epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science. It studies a natural phenomenon, viz., a physical human subject. This human subject is accorded a certain experimentally controlled input --

²⁰⁶ Psychological naturalism is a sort of “naturalism” because it posits that “human beings are . . . subject to . . . laws of nature,” that is, to “laws of some or other natural science;” but it need be neither reductionist nor materialist because one can defend the “scientific status of . . . psychology directly, without seeking any sort of reduction” (Botterill and Carruthers, pp. 1, 186-7).

certain patterns of irradiation in assorted frequencies, for instance -- and in the fullness of time the subject delivers as output a description of the three-dimensional external world and its history. The relation between the meager input and the torrential output is a relation that we are prompted to study for somewhat the same reasons that always prompted epistemology: namely, in order to see how evidence relates to theory, and in what ways one's theory of nature transcends any available evidence...But a conspicuous difference between old epistemology and the epistemological enterprise in this new psychological setting is that we can now make free use of empirical psychology. (Quine, 1969: 82-3)

As critics have noted (see, e.g. Kim 1988: 390), such an approach risks giving up the normative dimension of epistemology. As we saw with respect to evolutionary naturalism, there is no reason to think that the theories of natural sciences will be able to make sense of the normative foundations of knowledge claims as such.

In practice, therefore, most attempts at naturalism in epistemology today are more modest than Quine's.²⁰⁷

A fundamental goal of psychology is to describe how humans reason. A fundamental goal of epistemology – the theory of knowledge – is to set out how humans ought to reason, and so to acquire knowledge. There is no responsible way of answering the second question without accurately answering the first.²⁰⁸

Even at this level of generality, Kant would resist epistemic naturalism. In general, his response to psychological naturalism will be similar to his response to evolutionary and neuroscientific naturalism. For the purposes of studying human beings as empirical objects, naturalism is appropriate, so the best scientific psychology should give the most empirically adequate characterizations of and explanations for humans' thoughts, feelings, motives, and behavior. But naturalism cannot adequately make sense of the normative demands implicit within humans' transcendental standpoint. Rather than recapitulate this argument in general terms, the rest of this section focuses on two recent philosophical attempts to appropriate insights from empirical psychology to make normative claims about human beings. Showing how Kant might respond to these attempts will further highlight the important distinction (and relationship) between transcendental and empirical anthropology.

²⁰⁷ Even Quine's own naturalism has grown more modest. Cf. Quine 1990.

²⁰⁸ J.D. Trout, from the course description of his course in epistemology, available online at <http://www.jdtrout.com/?q=node/35>, accessed 3-3-2009.

One arena of contemporary psychological research that has garnered substantial philosophical attention is the so-called “biases and heuristics” research program pioneered by Daniel Kahneman and Amos Tversky. This research program has shown a striking degree of irrationality in human thinking, even among experts thinking about highly significant but fairly straightforward problems in their field of expertise. It has highlighted several pervasive forms of human irrationality, including the “fundamental attribution error” (attributing behavior to character traits rather than situational factors), “base-rate neglect,” self-serving bias (the “Lake Wobegon effect”), and various illusions that arise from seeking to make our experiences match our expectations. For one example (base-rate neglect), faculty and students at Harvard Medical School were given the following problem:

If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5% [i.e., 5% of people who take the test falsely test positive for the disease], what is the chance that a person found to have a positive result actually has the disease . . .? (Casscells, Schoenberger, and Grayboys 1978, cited in Bishop and Trout 2005: 122)

Almost half of the respondents (and a much higher percentage of non-experts) answer that the chances of having the disease are 95%, and only one in five respondents give the correct answer (the chances are actually less than 2%).²⁰⁹ So far, all of this is only empirical description of how humans in fact reason. But this and similar studies of human rationality call into question our ability to make good decisions, and they have led some epistemologists to argue for radical revisions in how we ought to employ our reasoning ability. Michael Bishop and J.D. Trout use this research to offer a wholesale rejection of what they call “Standard Analytic Epistemology” (Bishop and Trout 2005: 104-118). Among other things, they argue “that it would often be much better if experts, when making high-stakes judgments, ignored most of the evidence, did not try to weigh that evidence, and didn’t try to make a judgment based on their long experience” (Bishop and Trout 2005: 25). Because of the unreliability of basic cognitive processes, we should replace those processes with others that are empirically demonstrated to be more reliable.

²⁰⁹ Most people reason that since the false-positive is 5%, the 95% of positive tests must be true. The correct answer is given by reasoning as follows: For every 1000 people, 50 (5%) will falsely test positive and one will truly test positive. Thus there will be 51 people who test positive, one of whom actually has the disease. That is, the chance that the person actually has the disease is slightly less than 2%.

Contemporary research in psychology is also used in ethics. One prominent use has been “situationist” critiques of character-based ethical theories.²¹⁰ Psychological research has increasingly shown the context-sensitivity of human decision-making, and philosophers like John Doris and Gilbert Harman use this research to critique character-based ethics: “The experimental record suggests that situational factors are often better predictors of behavior than personal factors To put it crudely, people typically lack character”(Doris 2002:2, cf. Doris 2002:5-6, 15-22 and Harman 2000: 168, 178).²¹¹ In one particularly dramatic example, students at Princeton seminary were invited to participate in a study of religious vocation. Subjects filled out questionnaires and were told to give a verbal presentation on the story of the Good Samaritan (Luke 10:25-37) in another building. After the questionnaire, subjects were told that they were either late, on time, or early for the presentation. Along the way, the subjects passed an (apparently) extremely distressed person. Whether students stopped to help correlated strongly with their level of hurry, with only 10% of the “high hurry” subjects stopping and 63% of the “low hurry” subjects stopping. In this and many other cases, circumstances are better predictors of behavior than character. Many philosophers take “situationist psychology” to imply that “[r]ather than striving to develop characters that will determine our behavior in ways substantially independent of circumstance, we should invest more of our energies attending to the features of our environment that influence behavioral outcomes” (Doris 2002: 146).

How would Kant respond to these and similar sorts of developments? Starting with the empirical psychology itself, Kant’s empirical anthropology is not only compatible with but actually anticipates the findings of both the biases and heuristics program and situationism. Most of its empirical detail of Kant’s account of cognition comes in his description of various “prejudices” that explain how people diverge from ideal ways of thinking. Like the biases and heuristics program, Kant both characterizes the effects of these prejudices and diagnoses their underlying grounds. Of course, the specific principles that Kant lays out are not the same as those discovered recently, and Kant’s introspective methodology differs substantially from the experimental and statistical methods being used today. But the overall

²¹⁰ Although these criticisms are typically directed against theories of “virtue ethics” inspired by Aristotle, they seemingly apply as well to Kant’s own moral theory, within which action on the basis of consistent maxims that constitute one’s character is a central feature of moral decision-making.

²¹¹ For an excellent defense of ancient virtue ethics against these sorts of critiques, see Kamtekar 2003.

structure of Kant's account – supplementing a logic of ideal thought-processes with detailed empirical studies of systematic divergences from those ideals – is consistent with the biases and heuristics program. Similarly, with respect to human choice and action, Kant insists that character in the strict sense is “rare” (7:292); while most people have something like character, in that they often act on the basis of principles, Kant – like situationists – argues that which specific principles people actually act on depend, often in ways that they do not acknowledge themselves, on contingent circumstances and inclinations. Kant's empirical account of action requires amendment and refinement in the context of recent situationist research. More actions might be motivated by lower faculties that Kant envisioned, and the ways in which character is affected by circumstances seem to be more complicated than he supposed. But the fundamental structure of Kant's empirical anthropology is not challenged by this work.

With respect to both situationism and biases and heuristics, however, Kant's ability to endorse key findings of contemporary psychological research would be conjoined with vehement rejection of the dominant ways that naturalist philosophers make use of that research. This anti-naturalism is clearest in the context of Kant's ethics. Whereas ethical naturalists see in situationist psychology a reason to “invest . . . our energies attending to the features of our environment that influence behavioral outcomes” (Doris 2002:146), Kant sees situationist psychology as an empirical confirmation of humans' radical evil (see Frierson 2010c). Rather than accomodating the demands of morality to the general lack of character among human beings, Kant argues for precisely the opposite emphasis. Since Kant gives good a priori grounds for the moral importance of character,²¹² the rarity of character provides a reason to do empirical research on the means by which character can be cultivated. Kant suggests specific, empirically-informed techniques for cultivating character in oneself and others, including such things as the importance of avoiding even apparently innocent dissembling, keeping company with specific sorts of people, and “moderat[ing] our fear of offending against fashion” (7:294). As with much of his pragmatic anthropology, these suggestions are based on limited empirical knowledge. But rather than subordinating moral philosophy to situationist psychology, Kant's approach suggests the importance of devoting resources to studying the means to cultivating and fostering character. Thus rather than using the fact that situational variables were highly explanatory of behavior in the Princeton seminary case, Kant would focus on the 10% of “high hurry”

²¹² See chapters one and three, and Frierson 2006.

subjects that did stop, in order to gain insights that might make it possible to better foster strong character in a larger range of people. If all that matters morally is maximizing good behavior or consequences, then one might reasonably take situationism in psychology to imply that resources should be devoted to putting people in situations conducive to behaving well. But if, as Kant argues, it matters morally whether or not one acts from a good character, then one cannot ignore character even if it is difficult to cultivate.

Similarly, with respect to epistemology, Kant can and does make use of empirical insights into flawed reasoning for the purposes of pragmatic anthropology. The fact that people err in reasoning in predictable and systematic ways gives one good reasons to develop techniques for counter-acting these errors and for cultivating reasoning abilities less liable to error. But the facts about how people do reason cannot in themselves set the standard for how people ought to reason. One example of such normative divergence can be highlighted by the recent attempt by Bishop and Trout to use empirical studies of human reasoning to justify reasoning strategies such as the use of “statistical prediction rules,” simple rules based on a few variables that highly correlate with desired predictions. For example, to figure out whether a married couple will be happy (or at least, report that they are happy), “take the couple’s rate of lovemaking and subtract from it their rate of fighting” (Bishop and Trout 2005: 30). Or, to figure out whether a particular patient is neurotic or psychotic, use the “Goldberg Rule,” a formula based on the patient’s MMPI (Minnesota Multiphasic Personality Inventory) profile. The point is not merely that these rules should be used as part of deliberation, but that one should take them to *trump* one’s own considered judgment, even if one is an expert. Thus no matter how well you think you know a couple, you would do better to use the simple “lovemaking-minus-fighting” rule than to judge based on your experience, and no matter how sophisticated a clinician you are, you should use the Goldberg Rule rather than your own careful and detailed assessment of the patient’s mental state. The evidence for this is that, quite simply, moderately good SPRs work better than even very good expert opinion: “when tested on a set of 861 patients, the Goldberg Rule had a [success] rate of 70%; clinicians’ . . . varied from . . . 55% to . . . 67%” (Bishop and Trout 2005: 89). Relative to judgments based on long experience and careful examination of all available evidence, simple prediction rules take less effort, require fewer facts as inputs, and give more accurate results.

Bishop and Trout recognize that as tidy as these rules look from outside the process of reasoning, it is very difficult to remain faithful to them in practice:

We understand the temptations of defection. We know what it's like to use a reasoning strategy of proven reliability when it seems to give an answer not warranted by the evidence. It feels like you're about to make an unnecessary error. And maybe you are. But in order to make fewer errors overall, we have to accept that we will sometimes make errors we could have corrected, errors that we recognized as errors but made them nonetheless . . . People often lack the discipline to adhere to a superior strategy that doesn't "feel" right. Reasoning in a way that "feels" wrong takes discipline. (B&T 2005: 91)

When thinking about a pair of close friends who don't make love much, who fight about substantive issues and end up "stronger for it," who work and play together and have kids that they love and care for, it can seem insane to limit my judgment about their happiness to a simple "lovemaking-minus-fighting" calculus. But Bishop and Trout claim, for a variety of cases like this one, that we precisely *should* ignore the additional experience and evidence that we think is relevant and focus just on the simple formula. From-within, such a strategy can seem irrational, and hence Bishop and Trout say that they "understand" the temptation to defect. But we should resist that temptation so that we will get better epistemic results.

Despite their assurances to the contrary, however, Bishop and Trout's emphasis on "feeling" suggests to me that they do not *really* understand the temptation to defect. The problem is not merely that something "feels" wrong, but that ignoring evidence violates an epistemic standard to which we hold ourselves from-within. Bishop and Trout suggest revising epistemic standards away from reasoning based on what seems to be good evidence towards reasoning that brings about good effects, whether alethic (truth-conducive) or otherwise: "The primary aim of epistemology, from our perspective, is to provide *useful*, general advice about reasoning" (Bishop and Trout 2005: 94, emphasis added), where advice is "useful" towards the end of "human well-being," including such things as "avoid[ing] pain and misery" (Bishop and Trout 2005: 94); "health, deep social attachments, personal security, the pursuit of significant projects" (Bishop and Trout 2005: 99), and even "discovering the truth about the basic physical or social structure of the world" (Bishop and Trout 2005: 97). This pragmatic purpose of knowledge might seem obvious, especially when we add discovering truth as one of the central aims of good reasoning. But in fact, it is not clear

that maximizing true beliefs about practically and theoretically relevant features of the world is or should be our highest epistemic value. Consider here Kant's famous slogan: "Sapere aude," or "think for yourself." Kant points out, precisely in the context of defending intellectual autonomy, that such thinking will, especially at first, lead to more error than simply trusting formulaic thoughts handed down by experts, and cautions that "the danger . . . makes [those starting to think for themselves] timid and usually frightens them away from any further attempt" (8:35). But for Kant, *autonomy* of thought is a value in itself. Kant does not spend much time *defending* autonomy as an epistemic value, but one might make an argument here akin to his argument for the value of moral autonomy. Human beings are cognitively free precisely because we have an ability to weigh evidence for ourselves. Insofar as we relinquish that capacity, we relinquish our freedom, and, in an important sense, no longer "think" at all.²¹³ Much more would need to be said here in order to fully defend what we might call a "deontological" epistemic standard against Bishop and Trout's basically consequentialist standard. But it should at least be clear that the application of psychological research on the alethic and practical benefits of certain reasoning strategies is not *sufficient* to justify those strategies epistemically.²¹⁴

So what should we do if we find that reasoning strategies that are most accurate and useful involve relinquishing epistemic autonomy? We can't complacently just accept that we will think false things when we try to think for ourselves, ignoring the data. And we also shouldn't abdicate our responsibility for thinking for ourselves. What is called for is a more sophisticated applied epistemology, one that looks for ways to maximize overall good thinking, which includes the value of autonomy. That is, we should focus on developing strategies and even rules of reasoning that do not require ignoring available evidence but that give effective tools for autonomously thinking through that evidence in ways that help us get accurate and helpful results while still genuinely thinking for ourselves. Kant, in his lectures on logic, offers some beginning of an empirical theory of helps and hindrances to good reasoning, and includes the importance (and a recognition of the difficulty) of using rules of reasoning

²¹³ We can see this if we think about how we might evaluate epistemic failure. If one weighs all the evidence and makes a judgment in accordance with one's best estimate of where the evidence falls, one's epistemic merit (or demerit) will be genuinely one's own. But if one simply applies a Statistical Prediction Rule, and that rule turns out to be misguided (even *systematically* misguided), it is not really *my* epistemic mistake. It is the mistake of the rule.

²¹⁴ To be fair, Bishop and Trout recognize this, and see "better identify[ing] what is involved in human well-being" as a necessary step for further research (Bishop and Trout 2005: 156). (Unfortunately, they mistakenly think that defining human well-being is largely an empirical matter (Bishop and Trout 2005: 99, 156).

consistently, diagnosing and unvieling prejudices, and incorporating social reasoning into individual judgment. Much more detail is needed, and a Kantian reorientation of the empirical discipline of ameliorative psychology might be just what is needed to facilitate greater accuracy and usefulness in reasoning while preserving that autonomy the loss of which “feels” wrong and “tempts” us to defect.

IV. Naturalism and Freedom

This chapter has only scratched the surface of the amazing progress in our empirical understanding of human beings over the past 200 years. From the standpoint of Kant’s empirical psychology, this progress is largely welcome. Kant fully accepted the possibility of natural-causal accounts of human cognition and activity. He developed psychological theories about the nature and functions of various human mental faculties and conjectured about the brain chemistry that made such mental faculties possible. And while Kant was skeptical of the possibilities for fully understanding either the brain-states that underlie human mental life or the historical causes of our basic mental structures, nothing about his philosophical approach to human beings precludes such developments. Kant’s transcendental philosophy even provides grounds for insisting upon a thoroughly naturalist account of human beings within empirical anthropology. At the same time, his pragmatic anthropology shows some ways of making use of empirical findings for improving human lives, and – like ameliorative psychology today – Kant insisted that empirical research on human beings should be put to practical use. But Kant’s transcendental anthropology suggests caution in taking developments in our empirical understanding of human beings to imply the sort of thoroughgoing naturalism that would preclude the need for a priori theorizing about human beings from-within. Empirical knowledge about how humans think, feel, or choose cannot establish the ultimate normative standards to which our thinking, feelings, and choices should be held. In that sense, Kant embeds a naturalist approach to science in modesty about science’s scope. Rather than seeing natural sciences as a God’s-eye perspective on all reality, Kant insists that they are human ways of understanding the world we experience. And these sciences must be supplemented with an account of human beings that makes sense of normativity in our lives. In laying out the relationship between transcendental and empirical anthropology, Kant’s philosophy not only solves a pressing problem of the modern age – how to take both science and values seriously – but

also helps cut off many of the more egregious misuses of contemporary scientific theories (to justify sloppy thinking or immoral actions).

So far, however, this chapter has side-stepped what is perhaps the most important point of contention between Kant's philosophy and contemporary sciences. Central to Kant's philosophy is his view that human beings are "transcendentally free," uncaused causes of changes in the world. Many have (rightly) seen that the sciences depend upon a more determinist conception of human beings and thus have (wrongly) taken the sciences to disprove Kant's account of freedom. Others have (rightly) recognized the importance of a Kantian conception of freedom for making sense of our lives and thus have (wrongly) tried to find a place for freedom within the natural sciences. Both of these approaches are (partially) misguided, but they reflect the real urgency of the problem of freedom. Echoing Kant (see Bxxix), the psychologist Steven Pinker has put the problem this way: "Either we dispense with all morality as an unscientific superstition, or we find a way to reconcile causation (genetic or otherwise) with responsibility and free will" (Pinker 1997:55).

Roughly speaking, we can outline four different ways of thinking about freedom in relation to contemporary natural science:

(1) Anti-normative fatalism. For many, this is the most natural response to insights of natural sciences into the brain-dependence of mental states, the genetic and memetic bases of human behavior, or psychological determinism. If the sciences can explain what a human being does by appealing to causes that are part of the natural world, and especially if these causes can in principle be traced to causes that pre-existed the birth of the human being, then human beings are not free and hence not responsible for our thoughts or actions (e.g. Pereboom 2001). (A rarely invoked variation on this theme is to declare science to be fatalistic but to reject science in favor of morality.)

There are several major problems with this view. First, it involves a non sequitur.²¹⁵ The fact that sciences can explain human behavior causally need not imply (as Kant showed in his Third Antinomy, see chapter one) that human beings are not free. Second, it overstates the result of contemporary human sciences. While scientists assume that human behavior can be explained in terms of natural causes, contemporary human sciences are far from succeeding in actually explaining human complexity in natural terms except in the broadest outlines. Finally, the view is *prima facie* self-undermining, at least when

²¹⁵ Actually, it involves several non sequiturs, but I focus on one.

applied to epistemic norms. If causal determination precludes normative evaluation, then the natural sciences themselves are merely successful memes, with no legitimate claim to truth.²¹⁶

(2) Indeterminism.²¹⁷ In fact, contemporary natural sciences, unlike the Newtonian science of Kant's day, do not think that everything in the universe is deterministically-caused; quantum mechanics postulates indeterminism in nature, and the complexity of the human brain can give rise to contexts within which quantum indeterminism can make a significant differences for human thought, choice, and behavior. Robert Kane has suggested that "physical modeling in the brain" that incorporates "neural network theory, nonlinear thermodynamics, chaos theory, and quantum physics" can "put . . . the free will issue into greater dialogue with developments in the sciences" and, in the end, provide for a scientifically-plausible view of free will that justifies "the power of agents to be the creators . . . of their own ends and purposes" (Kane 1996: 17, 4). Indeterminism in this sense fits with a strong scientific realism, in that the sciences can exhaust all that there is to know about human beings. Because the sciences themselves are indeterminist, however, some human behavior may be as well, and this leaves room for freedom in human life.

Unfortunately, this view confuses indeterminism with freedom. Unless I have reason to identify with quantum fluctuations in my brain rather than deterministic processes shaped by my genetic and environmental background, the "ends and purposes" arising from those quantum fluctuations will be no more my own than those arising from deterministic influences. Because self-image is largely shaped by features that are stable or at least consistent with our past personality, changes that arise from quantum fluctuation may even seem *less* my own than those that are strictly determined. Moreover, since the physical models by virtue of which human beings are free posit chaotically complex systems that are only sometimes affected in significant ways by quantum fluctuations, there is no way to know, for any particular end or purpose, whether that end or purpose is really free, which undermines much of

²¹⁶ One might, of course, be an *immoralist* fatalist – denying *moral* responsibility – while still clinging to epistemic norms in a compatibilist way and thus assuming the legitimacy of norm-governed freedom from-within the standpoint of cognition while denying its legitimacy from-within the standpoint of volition.

²¹⁷ There are two main views among philosophers who take this approach as to the role of indeterminism in free action. One view, which is known as "simple indeterminism," holds that an action is only free if it is not caused, or not deterministically caused, by prior events. The second view, which is known as "causal indeterminism," holds that free action must be indeterministically caused by the right kinds of events. For a defense of the first view, see Ginet 1990 and 1997. For a defense of causal indeterminism, see Kane 1996 and 2002.

the practical value of positing freedom. (It will not be the case, for example, that all cases where one would naturally ascribe moral responsibility will be cases within which the relevant quantum fluctuation was present.)

(3) Compatibilism. Compatibilists, like indeterminists, seek to find freedom within scientific accounts of human beings. But compatibilists do not aim to find room for freedom in the indeterminism of the natural world. Instead, compatibilists aim to show that mysterious and metaphysical “transcendental” freedom is not the sort of freedom that human beings need. What is needed for making sense of moral responsibility, normativity more generally, and even just our sense of ourselves as free is something much more mundane, a sort of ability to impact what happens in our lives and our world. If this ability is itself grounded in genes or brain-states or psychological structures, that is not particularly important. What matters is that it is an ability that we can identify with and that we can see as genuinely efficacious in the world. Some sorts of determinism might make it hard or impossible to identify with aspects of our psychological make-up. If we recognize that our inability to focus on our work is genetically programmed or that our fear of spiders is a childhood phobia, we might not think of those aspects of our psychology as really “us.” But the mere fact that some aspect of our psychology is determined need not preclude us from identifying with it.

Compatibilism might provide an adequate conception of freedom, and recent philosophical work on freedom has provided substantial resources for conceptions of freedom that could fit within a wholly naturalistic approach to human beings. In his own transcendental analyses, especially of moral responsibility, Kant argued that morality requires a transcendental freedom that stands above any determination by natural causes.²¹⁸ Recent years have seen many attempts to make sense of moral responsibility without assuming transcendental freedom. Such attempts, however, cannot be “naturalist” in the sense of merely “clarify[ing] and unify[ing]” scientific theories (Dennett 2003:13). Rather than starting with science, figuring out what sort of freedom it allows for, and then arguing that this freedom is sufficient, a compatibilism that would do justice to our from-within sense of moral responsibility must start from-within, look carefully at the presuppositions of our conception

²¹⁸ See chapter one, pp. xxx-xxx.

of moral responsibility, and then see whether this is compatible with science.²¹⁹

(4) Perspectivism. The previous three approaches to freedom all fit well with strong scientific realism.²²⁰ Throughout this chapter, however, I have emphasized that Kant's contribution to debates about the natural sciences is his perspectivism. By recognizing that science represents one perspective on the world, Kant is able to make room for other perspectives, including a practical perspective within which freedom plays an important role. Many contemporary natural scientists have adopted a similar view. Dawkins and Dennett point out that human beings are capable of taking a stance towards the world that is not reducible to their genes or memes. Steven Pinker has, more forcefully, insisted that "science and ethics are two self-contained systems played out among the same entities in the world" (Pinker 1997: 45). Perspectivists can thus defend the integrity of an incompatibilist conception of freedom as a concept that, as Pinker puts it, "makes the ethics game playable" (Pinker 1997: 45). For perspectivists of this stripe,²²¹ compatibilism is wrong in trying to find a notion of freedom that is compatible with our best scientific theories. Freedom is needed within ethics; causation is needed within science.

Importantly, one can prioritize perspectives in different ways. Most natural scientists and philosophers heavily influenced by natural science are what we might call science-first perspectivists. Pinker is typical here:

Ethical theory requires idealizations like free, sentient, rational, equivalent agents whose behavior is uncaused, and its conclusions can be sound and useful even though the world, as seen by science, does not really have uncaused events . . . [T]he world is close enough to the idealization of free will that moral theory can meaningfully be applied to it. (Pinker 1997: 55)²²²

²¹⁹ For recent attempts to do this, see Frankfurt 1998 and Wallace 1994 and 2006. Even many contemporary Kantians have backed off a strong commitment to transcendental freedom as a condition of possibility of moral responsibility. See Korsgaard 1996.

²²⁰ In all of these cases, of course, one need not be a strong scientific realist. Theological compatibilists, for instance, might argue that God is the ultimate cause of all things and causes them in accordance with natural laws, but that human freedom exists nonetheless. For the sake of comparing Kant with contemporary naturalism, though, I treat as strongly realist all of those views of freedom that are compatible with strong scientific realism.

²²¹ One could be a compatibilist perspectivist (as Dennett is) insofar as one takes the ethics game to require a conception of freedom that the science game need not, but also takes the conception of freedom in the ethics game to be entirely compatible with conceptions of the self that are evolutionarily explicable.

²²² Dennett's claim that "the creation of . . . new standpoints" is "the . . . product of . . . [a biological] revolution" (Dennett 2003: 179) also fits here.

In thinking of the world as “close enough” to ethical assumptions, Pinker implicitly assumes that the world “as seen by science” is the real world, and ethics is all right because its assumptions are not too far off, good enough for practical purposes. Pinker compares ethics to Euclidean geometry in this respect. But this sort of perspectivism cannot be adequate to make sense of the demands of normativity. If we are in fact only approximately free, then either the “ethics game” (Pinker’s term) requires only approximate freedom to be legitimate (as, for instance, getting good results in structural engineering requires only that the world be approximately Euclidean), or the ethics game, while playable, is a sham. Perhaps Pinker is willing to affirm some sort of idealized uncaused freedom merely in the absence of the well-worked-out compatibilist account that would show that the freedom we really need is actually compatible with science. But if morality really depends upon seeing ourselves as uncaused, it is just not clear how a determinist world can be “close enough” to save it.²²³

Instead of a science-first perspectivism, then, one might adopt a neutral perspectivism, as has become increasingly common amongst contemporary Kantians. Christine Korsgaard, for example, argues that the fact that “freedom . . . is not a theoretical property which can . . . be seen by scientists” will be taken to imply “that . . . freedom is not ‘real’ only if you have defined ‘real’ as what can be identified by scientists looking at things . . . from outside” (Korsgaard 1996b:96). But there is no reason to do this, since “we need” a from-within practical perspective – and thus freedom – just as much as we need scientific theories. This approach is considerably more promising because it preserves all of the insights of science without according science an unjustifiably privileged place in our self-understanding. And it thereby avoids the most serious problems of science-first perspectivism.

But Kant (and some contemporary Kantians) offer good reasons to reject even neutral perspectivism in favor of freedom-first perspectivism. On this view, the scientific view of the world is subordinate to the view of the world according to which human beings are free. The predominant argument for the priority of this free perspective on human beings within

²²³ Here it’s important to remember that morality is meant to be saved in the sense of *justified*. Of course the world can be set up to save morality in a *psychological* sense, that is, as a game that human beings can and will play. But this does not show that morality is a *legitimate* game, one that we *should* play. Moreover, if freedom is necessary not merely for ethics but for normativity in general, then it seems at least odd to see the conditions of possibility of science itself (freedom) being subordinated to the best theories within that science. Especially if these theories posit that we are unfree, they might be able to show that the practice of theory-construction in science is something that we will in fact do, but it would undermine the basis for thinking that our scientific theories are really *justified*.

contemporary philosophy is simply that any perspective that claims any sort of justification – even the scientific perspective itself – implicitly appeals at least to the freedom to believe on the basis of normative standards of good evidence. In holding each other and ourselves responsible for our beliefs, whether scientific or otherwise, we treat those beliefs as “up to us” rather than mere effects of empirical causes. But there is no corresponding dependence of normative perspectives upon scientific ones. Once one treats human beings as natural objects, one must explain the capacity for norm-governed thought and action in terms of science, but one can think of oneself and others as norm-governed without committing oneself to a scientific picture of the world.

Kant offers a further reason for freedom-first perspectivism,²²⁴ based on a fundamental difference between the “ethics game” and what we might call the “science game.” Humans treat each other as morally responsible, which depends upon seeing each other as free; and we study each other empirically, which depends upon seeing each other as determined by natural causes. So far, the games are parallel; each depends upon a certain assumption as its condition of possibility. But the parallel nature of these assumptions conceals a deeper difference between them. For the ethics game, as we saw above, it is not enough that we seem free, or that we are close enough to being free. In that case, we might still engage in the ethics game, but the game itself would be a sham. For the science game, however, it is not necessary that we really be wholly determined by natural causes. It is enough that human behavior, whatever its ultimate basis, is sufficiently regular to be explained in terms of natural causes.²²⁵ Since the science-game does not depend upon a science-first perspective and the ethics game does depend upon a freedom-first perspective, we can and should adopt the freedom-first perspective.

²²⁴ He offers at least one other “reason” that I do not discuss. Aesthetic experience, of the sublime, allows us to feel the priority of our freedom.

²²⁵ Moreover, as Kant argues in his *Critique of Pure Reason*, the demand for ultimate scientific explanations is impossible. As one contemporary commentator has put it:

all naturalistic explanations – even the most impressive explanations of some future neuroscience – are conditional explanations In a certain sense they are incomplete, for they can never explain that any natural law should take the form that it does. (O’Neill 1989: 68)

Science at most explains regularities in nature in terms of increasingly general laws, but these explanations are always incomplete in a way that leaves room for the sort of incompatibilist compatibilism Kant defends in his transcendental philosophy. Practically speaking, since the evidence that some empirical cause will bring about an action of ours can never ultimately be based in a causal law that is self-evidently necessary, it always remains open to us to make an exception of ourselves.

V. Conclusion.

In the end, where does modern science leave us with respect to Kant's question? We now have a much more sophisticated empirical understanding of human beings, from our complex psychology to the brain-chemistry that makes this psychology possible to the evolutionary origins and genetic bases of our wonderful brains. Kantians can and should embrace the results of the natural sciences into empirical anthropology. And that means, of course, that Kantians can and should embrace the methodological naturalism that makes these sciences possible; humans should be treated as ordinary objects in the natural world and studied according to the best methods of natural science. Kant even provides grounds for embracing a modest scientific realism, in which one takes the methodological assumptions of natural science to be empirically real, that is, to be constitutive of the world we experience.

As we have frequently noted, however, the question "What is the human being?" is not merely a question about the distinctive features of a certain type of natural entity. Answering the question requires thinking not only about how to pick out homo sapiens from amongst other natural objects, but about how to make sense of ourselves, from-within. And this question from-within is not merely philosophical but also practical, a matter of asking what to do with our lives, what to think, and what to find pleasure in. For those questions, naturalism is insufficient. But Kant's philosophical framework not only includes an empirical realism according to which science is true of the empirical world, but also a transcendental idealism that insists that science is only one perspective, that there is more to the world-in-itself than what is captured in the empirically-knowable world. In Pinker's terms, Kant allows for both the science-game and the ethics-game, or, more generally, the normativity-game. Kant's justification of the science-game allows for a realism sufficiently robust to allow us to pursue and benefit from science, but it mitigates that realism just enough to make room for freedom. Moreover, precisely because Kant engages in both the science-game of empirically studying human beings and the normativity games of epistemology, ethics, and aesthetics, he provides a better model for answering the question "what is the human being?" than modern naturalism. Kant justifies the fundamental normative standards of epistemology, ethics, and aesthetics independent of the natural sciences; but he is still able to use the results of those natural sciences in the context of a pragmatic

anthropology that thinks about how we, as empirically-knowable human beings, can best promote ways of thinking, acting, and feeling that conform to our ideals.

Of course, the fact that science must be complemented by a perspective from-within does not show that Kant best analyzed that perspective. Just as modern science accepts the importance of empirically studying human beings while moving beyond Kant's specific empirical theories, one might accept that normativity is not reducible to empirical science and still reject Kant's specific normative theories. And in fact, philosophy in the past two centuries has developed new ways of thinking about human beings from-within that directly challenge Kant's transcendental anthropology. In chapter ten, we look at one of the twentieth century's defining "from-within" approaches to human beings: existentialism. And in chapter eleven, we survey some dominant contemporary philosophical approaches to normativity. Before turning to those, however, we turn to a further sort of empirical anthropology. Like contemporary biology and psychology, Kant typically focused on those aspects of human beings that were, at least in a loose sense, universal. But Kant also recognized that human beings are historical in ways that no other animals are, and he recognized that the human species includes within it sub-groups with their own distinctive characteristics. The next chapter turns to contemporary developments that emphasize historical and cultural differences between human beings. Kant saw these differences as primarily a small sub-field within empirical anthropology, but recent thinkers have extended historicism and diversity into transcendental anthropology. This sort of historicism goes far beyond Kant's own and raises a substantial challenge to his anthropology as a whole.

Chapter 9: Historicism and Human Diversity

Perhaps the time is at hand when it will be comprehended again and again how --little used to be sufficient to furnish the cornerstone for such sublime and unconditional philosophers' edifices as the dogmatists have built so far: any old popular superstition from time immemorial (like the soul superstition, which, in the form of the subject and ego superstition, has not even yet ceased to do mischief); some play on words, perhaps, a seduction by grammar, or an audacious generalization of some very narrow, very personal, very human, all too human, facts.

—Friedrich Nietzsche, *Beyond Good and Evil* (Preface)

It seemed to me that, for the moment, the essential task was to free the history of thought from . . . transcendental narcissism.

—Michel Foucault, *The Archaeology of Knowledge* (Foucault 1982: 203)

Kant's lifetime saw a rise in consciousness of human beings as historical beings. Throughout the Enlightenment, philosophers and scientists had in one sense seen themselves in historical terms; the Enlightenment was a period of rejecting the contingent, tradition-bound ideologies of the Middle Ages. But the alternative to these traditions was not taken to be just another tradition. Instead, Enlightenment philosophers saw themselves as replacing merely historical traditions with ahistorical truths grounded in reason and direct experience of the world. In one sense, Kant too is an Enlightenment thinker. His transcendental anthropology aims to lay out the a priori – and hence necessary and universal – structures of human thought and action, stripped free of anything merely empirical (including anything merely historical). At the same time, as we saw in chapter three, Kant was attuned to fact that human beings are historical; he even emphasized the historicity of Enlightenment itself.²²⁶ Kant's students (Herder) and followers (Reinhold) further emphasized this point. By the time of Hegel and Marx, the idea that humans' fundamental ways of thinking about and acting within the world differ from culture to culture and change

²²⁶ See Foucault's discussion of Kant's "What is Enlightenment?" (8:33-42) in Foucault 1984: 32-50.

from one historical epoch to another had become commonplace. Today, this emphasis on human variety pervades not only history but also sociology, anthropology, cultural studies, and even parts of psychology and literary theory.

But human “historicity” is ambiguous. In its most mundane sense, the claim that human beings are historical is one that (virtually) no one would deny. Humans age. We change as we age. We live in communities that develop cultural practices. These practices get passed on and modified; they vary from time to time and culture to culture. But historicism is generally not limited to this mundane historicity. In chapters six and seven, we saw that two importantly different sorts of historicism emerged in the nineteenth century. One, represented by philosophers such as Hegel and Marx, emphasized the necessity of historical change. For Hegel, historical changes chart rational progress. For Marx, material changes in humans’ labor and power relations give rise to both ideological shifts and the need for further historical shifts. Other philosophers, such as Herder, Hamaan, and Nietzsche, use historicity and diversity to emphasize the contingency of particular modes of thought and forms of life. For both sets of historicists, humans’ historical nature is not limited to external studies of changing individuals and cultures. Over history and across cultures, human beings differ in the ways they see the world. In Kantian terms, transcendental anthropology is historicized and localized. No one can elucidate conditions of possibility of human cognition or volition; instead, as Nietzsche put it, those who “wanted to provide a rational foundation for morality” are exposed as having only justified “the morality of their environment, their class, their church, the spirit of their time, their climate and part of the world” (Nietzsche 1966:§186). In the rest of this chapter, I use the term “historicism” to refer to this stronger sort of historicism, what we might call a “transcendental historicism” since it claims that humans’ from-within, normative perspectives are historically conditioned.²²⁷

This historicism has become a mainstay of our world. In its most facile form, it manifests itself in a lame relativism that only refers to what is true-for-me or good-to-me. Such relativists generally think that the mere diversity of human ways of thinking establishes that no particular

²²⁷ While Hegel, Marx, and Nietzsche all fall into this strong sense of historicism because all three emphasize the historicity of what Kant considered transcendental anthropology, the figures on which I focus in this chapter (Kuhn, Foucault, and contemporary anthropologists) are more specifically Nietzschean, in that all emphasize (to varying degrees) the contingency of historically varying perspectives on the world. (In chapter 11, we will look at a more Hegelian approach to history in the work of Alasdair MacIntyre.)

perspective on the world can be right. And such relativism is often confused with tolerance, as though the most respectful stance to take towards different times and cultures is to recognize that their beliefs and values were true for them, just as one's own beliefs and values are true for oneself. At the end of this chapter, I will come back to this simplistic but all-too-common appropriation of historicism, but for the bulk of this chapter, I focus on three significant and more nuanced recent versions of historicism.

I start with the rise of historicism in the study of the natural sciences, focusing on Thomas Kuhn, through whom the concepts of a “paradigm” and a “paradigm-shift” have become commonplace. Kuhn and post-Kuhnian philosophy of science is important because it seems to undermine not only strong scientific realism but even Kant's own modest realism about natural science. If natural sciences are historically conditioned, it becomes hard to see how one can even talk about the empirical world or take Kantian categories of experience to have the strict universality Kant thinks they do. From Kuhn, I turn to Foucault. Like Kuhn, Foucault questions basic aspects of Kant's transcendental framework, but Foucault focuses on the problematic and historically contingent notion of “the human being” as such. In particular, while Kuhn calls into question Kant's transcendental anthropology of cognition, Foucault historicizes both the general framework of Kantian anthropology – what Foucault calls “man” as an “empirico-transcendental doublet” (Foucault 1970: 319) – and the conception of human agency that underlies Kant's moral theory. Finally, I turn to anthropology²²⁸ and cultural studies, where human diversity – rather than historicism as such – provides a perspective within which the supposed universality of Kant's anthropology (both empirical and transcendental) is questioned.

I. Historicism and Contemporary Natural Science: from Kant to Kuhn

The roots of historicism in the natural sciences lay in mid-19th century developments within mathematics, Kant's paradigmatic candidate for a priori knowledge. Mathematicians such as Nikolai Lobachevsky, Henri Poincare, and Bernhard Reimann began thinking of basic geometrical axioms – such as that parallel lines never meet – not as

²²⁸ Since this book has generally used the term “anthropology” to refer to the study of human beings in general, it is worth noting that in this chapter, I often use anthropology (as here) to refer to the narrow discipline of contemporary anthropology.

intuitively obvious truths about space but as “conventions” or logical axioms. Mathematicians began working through implications of changing various Euclidean axioms, and a whole field of non-Euclidian geometry was born. In these systems, familiar geometrical claims – such as that the sum of the interior angles of a triangle is 180 degrees – no longer hold true. In itself, this development in mathematics was important since it suggested that the human mind was capable of thinking about worlds with structures very different from our own. But Euclid’s axioms were still generally taken to define the true nature of space. Mathematicians might have fun thinking about a world where parallel lines touch and where someone drawing a straight line would end up drawing a line that, while always straight, nonetheless kept criss-crossing itself. But these mathematical fantasies were just that: fantasies.

Then Einstein argued that the world itself was non-Euclidean. Straight lines can cross themselves, parallel lines touch, and the interior angles of a triangle are not 180 degrees. At the same time, other radical changes were happening in the physical sciences. Quantum mechanics challenged basic notions such as the continuity of time, the determinacy of space, and the principle of causation. Within quantum mechanics, time no longer passes in a continual stream but in little jumps, or quanta. Objects are no longer located in particular spaces but are smeared out in waves of probability. And events in the world are not universally explicable in terms of causes and effects. Some things happen, literally, by chance. Not only was Euclid’s space replaced by Einstein’s, but Newton’s deterministic world was replaced rolls of the dice.

As philosophers increasingly sought to come to terms with these and similar developments in science, the Kantian model of science as built on a set of a priori synthetic claims about the nature of any possible experience seemed increasingly difficult to maintain. In a now classic paper (Quine 1951), W.V.O. Quine argued against the distinction between synthetic and analytic claims and the possibility of any knowledge that could not be changed in the light of experience. If Euclidean geometry could be abandoned for the sake of relativity theory and universal determinism abandoned for the sake of quantum mechanics, then “no statement is immune to revision”; any claim is open to empirical challenge, including “even . . . logical law[s]” such as the law of the excluded middle or the principle of non-contradiction (Quine 1951:40). In the end, Quine claims, “The totality of our so-called knowledge or beliefs, from the most casual matters of geography and history to the profoundest laws of atomic physics or even of pure mathematics and logic, is a man-made fabric which impinges on

experience only along the edges,” such that while experience can lead to revisions in our web of belief, “No particular experiences are linked with any particular statements in the interior of the field, except indirectly through considerations of equilibrium affecting the field as a whole” (Quine 1951:40).

Quine’s account of knowledge set the stage for Thomas Kuhn’s historical turn in the philosophy of science. Kuhn’s *Structure of Scientific Revolutions* rejected what had become dominant models of historical progress in science and set the agenda for much history and philosophy of science today. Most science textbooks present a history of science within which old illusions give rise to more and more accurate scientific theories. The scientific method is described as a method whereby one measures theories or hypotheses against empirical evidence, rejecting theories that fail to be confirmed by evidence. Over time, theories become better and better confirmed and thereby more and more reliable. Scientific revolutions occur when empirical evidence contradicts long-standing theories and these are replaced by more adequate theories. Thus the transition from Newtonian physics to the physics of Einstein and quantum mechanics is explained by empirical evidence (such as the perihelion of Mercury and the bending of light around the sun evident in the famous eclipse of 1919) that contradicted Newton’s theory. Against this conception of science and its history, Kuhn argues both that the “scientific method” does not consist in the attempt to falsify hypotheses and the rejection of those that fail to measure up to empirical data and that “scientific revolutions” are not unambiguous forms of progress.

Kuhn’s account of science involves a fundamental distinction between what he calls “normal science,” “extraordinary science,” and “scientific revolutions.” Normal science is what virtually all practicing scientists engage in virtually all of the time. It consists in “puzzle-solving.” This science takes place in the context of a “paradigm,” a “constellation of beliefs, values, techniques, and so on shared by the members of a given [scientific] community” (Kuhn 1996: 174) that supplies “a criterion for choosing problems that . . . can be assumed to have solutions” (Kuhn 1996: 37). The paradigm resists falsification by empirical data. Scientists often are unable even to see data that contradicts the paradigm – the paradigm structures “the perceptual process itself” (Kuhn 1996: 62).²²⁹ Even if scientists perceive data that

²²⁹ Kuhn points to experiments within which subjects were shown playing cards, some of which were abnormal, such as “a red six of spades.” When shown these cards and asked to identify them, “the anomalous cards were almost always identified, without apparent hesitation or puzzlement, as normal”

does not fit their paradigm, they initially interpret such data as due to personal failures to properly carry out their experiments. If such data are reproduced consistently, scientists will work to explain them in ways that preserve the paradigm itself. Finally, if they are unable to explain the empirical data, scientists will generally move on to other areas of research rather than reject the theories that the data contradict. As Kuhn put it, “Paradigms are not corrigible by normal science at all” (Kuhn 1996: 122).

Sometimes, however, empirical findings that conflict with scientific paradigms – “anomalies” – are sufficiently disturbing to provoke “extraordinary science,” examining the anomaly in greater detail and seeking to explain it within the broad contours of one’s paradigm. Generally, anomalies are resolved within the dominant paradigm with only minor tweaks. Sometimes, however, an anomaly is sufficiently problematic, or an alternative paradigm sufficiently attractive, to prompt a crisis and eventually even a “scientific revolution.” This transition from an old paradigm to a new one “is far from a cumulative process” (Kuhn 1996: 86), more like a “gestalt switch” than a refinement of the old paradigm. One comes to see science in a new light; theories and even data of the old paradigm are often not even translatable into the new one. Kuhn point out, for example, that whereas many people today think of Einstein’s physics as a refinement of Newton’s, within which Newtonian physics is merely an approximation, in fact “Einstein’s theory can be accepted only with the recognition that Newton’s was wrong” (Kuhn 1996: 100) and even the most basic “variables and parameters” in each’s theory – the variables referring to “time, mass, etc.” – have different meanings, different physical referents, in the two theories. The “fundamental structural elements of which the universe . . . is composed” are different in the two systems; the apparent similarity of Einstein’s laws of motion at slow speeds to Newton’s laws is merely superficial (Kuhn 1996: 101-2). Similarly, in any true scientific revolution, the whole “conceptual network through which scientists view the world” changes (Kuhn 1996: 102). Put another way, one at once deeply Kantian and deeply historicist, “after a revolution, scientists are responding to a different world,” or, even more radically, “when paradigms change, the world itself changes with them” (Kuhn 1996: 111).²³⁰

(Kuhn 1996: 63, see Bruner and Postman 1949 for the original study). People literally *saw* cards that matched their categories. Similarly, scientists literally see data in terms of the paradigm within which they work.

²³⁰ This historicism about the world permeates Kuhn’s work, though there are some glimmers of a more transhistorical conception of reality, as when he suggests that before a revolution is possible, “Nature itself

Scientific revolutions do not simply replace a falsified theory with one that fits all the empirical evidence. Often, in fact, new paradigms fit available evidence worse than old ones. Because previous evidence was collected in order to confirm the old paradigm, the new paradigm often has a very difficult time making sense of it. Kuhn points out that Copernicus's astronomy did not predict the motions of the planets any better than the Ptolemaic astronomy that it replaced, Lavoisier's chemical revolution that paved the way for modern chemistry "deprive[ed] chemistry of some actual and much potential explanatory power" (Kuhn 1996: 107, cf. 131), and "in this century the striking quantitative success of both Plank's radiation law and the Bohr atom quickly persuaded many physicists to adopt them even though, viewing physical science as a whole, these contributions created many more problems than they solved" (Kuhn 1996: 154). What new paradigms bring is not primarily a better way of dealing with old evidence but a whole new way of looking at the world, new criteria for success in science, new assumptions about what sorts of empirical problems are worth investigating, new ways of interpreting empirical data, new experimental techniques, and even new criteria for deciding amongst competing theories. Whereas "progress" makes sense "during periods of normal science" where a paradigm provides accepted standards of progress, it does not make sense as a way of measuring the shift from one paradigm to another. Thus, Kuhn claims, "a decision [between competing paradigms] can only be made on faith," or, more precisely, "There must be a basis, though it need be neither rational nor ultimately correct, for faith in the particular [paradigm] chosen" (Kuhn 1996: 158). But then "[w]e may . . . have to relinquish the notion, explicit or implicit, that changes of paradigm carry scientists and those who learn from them close and closer to the truth" (Kuhn 1996: 170).²³¹

Kuhn developed his notion of paradigm shifts primarily within the philosophy of natural sciences, and his overall theory has led to a wide growth in investigations of the historical and social conditions that shape scientific practice and theories. Thus sociologists of science have offered detailed studies of the social, cultural, and even psychological factors that shape developments of scientific theory and practice. Feminist philosophers of science have shown ways in which the male-dominance of scientific work has skewed the way that data is perceived, collected, and interpreted (e.g. Longino 1990). At their most extreme, historicist

must first undermine professional security by making prior achievements seem problematic" (Kuhn 1996: 169).

²³¹ See, however, Kuhn 1977, where he seeks to recover a sense of objectivity and even "progress" in science.

studies defend the social construction of “scientific facts” or “quarks” (see Latour and Woolgar 1986 and Pickering 1984). Kuhn’s concept of a “paradigm shift” has also filtered into our culture more generally, resulting in a historicist conception of knowledge according to which human experiences, beliefs, and practices are shaped by paradigms that structure not only how we interpret our world but literally what we perceive. Our world is constituted by our paradigms, and paradigms shift. They can differ between different groups of people and can change – often quite abruptly – over time. What the world looks like from-within, including the most basic norms for how one evaluates one’s beliefs, is historically-conditioned.

This Kuhnian historicist account of human knowledge is, in some respects, strikingly Kantian. Like Kant, Kuhn suggests that human experience is structured by forms of cognition that precede that experience. With his claim that different scientists inhabit different worlds, Kuhn even endorses something like Kant’s Copernican turn (see Kuhn 1966: 110). With Kant, moreover, Kuhn gives reasons to reject a strong scientific realism by showing how science results from presuppositions human beings bring to analysis of the world. Scientific theories, whether about atoms or about human nature, are always constructed in the context of prior commitments of those who construct them. While this need not wholly undermine the “rationality” of science nor even some sort of scientific “objectivity” (see Kuhn 1977), it calls into question the strong scientific realism implicit in philosophical naturalism. Kuhn goes further in this respect than Kant, claiming that “the notion of a match between the ontology of a theory and its ‘real’ counterpart in nature . . . seems . . . illusive in principle” (Kuhn 1966: 206, cf. too McMullin 1995).

But while he supports something like Kant’s Copernican turn and even something like the rejection of strong scientific realism implied by Kant’s transcendental idealism, Kuhn’s picture is radically different from Kant’s in one important respect.²³² Kant’s transcendental anthropology of

²³² In my discussion of Kuhn’s importance for Kant, I ignore the fact, obvious in Kuhnian hindsight, that Kant’s own empirical psychology reflects contingent paradigms of thought. This point would not be particularly problematic for Kant. Though proud of his tri-partite conception of the human mind, Kant recognized that his empirical anthropology was revisable. And even if not particularly attuned to how certain notions (such as that of a “faculty”) were historically contingent, Kant was well aware that central aspects of psychology in his day – including the use of physiognomy, which Kant largely rejected, and the role of natural predispositions, which Kant defended – involved presuppositions that might be overturned.

cognition seeks universal conditions of possibility of any human experience. Kuhn's paradigms structure experience for particular groups of human beings at particular times in particular contexts. Kuhn historicizes and relativizes Kant's Copernican turn. Paradigms that are "a priori" in the sense that they structure one's experience of the world are not necessarily "a priori" in the stronger sense of being unrevisable in the light of further empirical research (or even, as has been emphasized by post-Kuhnian sociology of science, of changing social conditions). What scientific theorizing looks like from-within is determined in part by historically contingent facts. This raises the prospect that the transcendental forms of cognition that Kant argues to be conditions of possibility of any human experience might merely be historically local paradigms. For Kant, the idea of data that could contradict transcendental forms of intuition or categories of experience was literally not humanly thinkable. For Kuhn, such data is not only thinkable but actual. The history of science shows that even the most apparently necessary claims can be abandoned in the context of a scientific revolution that radically restructures our whole approach to our world.

As a vague and general point, the assertion of historical contingency human cognition need not be fatal to Kant's transcendental anthropology. Kant's account of empirical concepts leaves room for the development of concepts that shape one's experience of the world, even when these concepts are themselves ultimately rooted not in the necessary structure of human cognition as such but in the contingent ways in which human beings have responded to particular sets of experiences. Kant's accounts of prejudice further suggest a framework for thinking about cognitive structures that are contingent but nonetheless "a priori" in the sense that they shape the way we experience the world. Kant's theory of biology comes even closer to a sort of Kuhnian approach to paradigms; in the case of natural organisms, experience of the world gives rise to an a priori principle for further investigation of that world. And Kant's philosophy of history provides some basis for thinking that human perspectives on the world can change. Thus Kant could allow that, in addition to *a priori and universal structures of human cognition*, there are also historically contingent mental structures that shape our experience of the world.

Unfortunately, historicist philosophy of science does not let Kant get off that easily. First, the details of Kuhn's history of science suggest that the particular structures Kant assumed to be a priori are not. For Kant, space, time, causation, and a continuum of degrees of sensible properties are among the most fundamental a priori conditions of the possibility of any human experience. By virtue of the aprioricity of space,

we can know that (Euclidean) geometry applies to the empirical world. By virtue of the principle of causation, we can know a priori that every alteration has a cause that determined it to occur. But relativity theory seems to require rejecting Euclidean space, and Einstein's notion of "space-time" is inconsistent with Kant's careful distinction between space and time. Meanwhile, dominant interpretations of quantum mechanics imply that deterministic causation is not universal, that alterations occur probabilistically, the result, at least in part, of random chance. Thus the specific positive metaphysical claims of Kant's transcendental anthropology of cognition seem to be rejected by the best science of our day.

But the problem is even deeper. Precisely because Kant's transcendental anthropology of cognition did such a good job picking out the best candidates for the most basic presuppositions of human experience, if even these presuppositions are historically conditioned, there seems little hope for *any* truly universal structure of human cognition. All categories by means of which we make sense of the world seem open to revision. The fact that some philosophers suggest rejecting the principle of identity ($a=a$) and even the principle of non-contradiction in order to better make sense of contemporary physics drives this point home even more forcefully. Not only the details of Kant's transcendental anthropology but even the very idea that there could be a universally-human transcendental structure of cognition seem vulnerable to historicist critique.

II. Historicism and the Human Sciences: Foucault

While the history of natural science contributes important historicist dimensions to understanding human cognition, an even more radical historicism has emerged in those sciences devoted to studying human beings as such. The hero of this brand of historicism is Michel Foucault, whose detailed historical analyses of key concepts and practices employed in human self-understanding threaten the universality of not only Kantian cognitive categories but his whole transcendental anthropology. By historically analyzing ways of thinking, Foucault challenges Kant's universalism, and the central target of Foucault's work – insofar as one can pick out a central target – is the historically-conditioned nature of human subjectivity itself. Foucault aims to show the historical emergence and thereby contingency of precisely the conception of the human being that lies at the heart of Kant's anthropology.

In developing his historicist approach, Foucault avoids blanket theoretical claims about human historicity. Claiming that “human thinking is always historically-bound,” like trite relativist claims that “all truths are relative,” is a self-undermining proposition, a purportedly absolute truth that all truths are merely relative. But Foucault neither assumes an omniscient posture nor makes such overarching pronouncements. Instead, while recognizing and even embracing the historical-situatedness of his own work, Foucault “analyzes specific rationalities” (Foucault 1982: 210), studying particular developments in structures of human knowledge and society. Thus Foucault’s first major work, *A History of Madness* (1961), traces the origin of our concept of “mental illness,” showing how “mental disease, with the meanings we now give it, is made possible” (Foucault 1988:270, Foucault 1961/2006:504) And Foucault’s *History of Sexuality* shows, among other things, how sexual categories and even the basic structure of ethical life shifts from ancient Greece to Christian Europe to the present. By emphasizing detailed studies of particular cases – the “gray, meticulous, and patiently documentary . . . accumulation of source material” (Foucault 1984:76) – Foucault eschews appeals to timeless truths without making absolute claims about the impossibility of such truths. Foucault’s detailed studies model historicist thinking while avoiding the self-contradictions of dogmatic theoretical relativism. Thus Foucault’s threat to Kant is not the threat of a theory that undermines Kant’s anthropology, but a historicist way of thinking that provides an alternative to Kant’s anthropology and shows that anthropology to be historically-local and historically-determined rather than universal.

For Foucault, the “accumulation of source material” is neither a way of tracing the factual flow of history nor a way of describing historical “progress” towards the present. Instead, like Kuhn,²³³ Foucault emphasizes the historicity of the basic structures of human thought and action. Foucault describes his approach as “deliberately both historical and critical, in that it is concerned . . . with determining the conditions of possibility of” particular forms of experience (Foucault 1994: xix). Foucault’s historical method includes two key components: “archeology” and “genealogy.” The former lays out the “episteme” of a particular historical epoch, the “epistemological field” that defines the “conditions of possibility” of knowledge at a particular time. Foucault refers to this as a “historical a priori” that is “not a condition of validity . . ., but a condition for the reality of statements” (Foucault 1969/1982: 127). Whereas Kant’s a priori categories of experience are purportedly universal, formal

²³³ The similarity to Kuhn is not wholly coincidental. Foucault’s early mentors included Bachelard and Canguilhem, whose histories of natural science anticipated Kuhn’s. See Gutting 1989 and 2001.

structures of any possible human cognition, Foucault's historical a priori is a historically contingent set of conditions that structure neither what could be thought nor what is actually justified but only what actually is thought at any given time. To this archaeological excavation of historical epistemological fields, Foucault adds a genealogical component that traces how different fields arise and change, appropriate and dominate one another. It is "the history of morals, ideals, and metaphysical concepts" (Foucault 1984: 86).²³⁴ Consistent with both Kuhnian historiography and Nietzschean genealogy, Foucault's histories emphasize contingency and complexity in historical changes and reject any sense of "progress" towards some supra-historical ideal (such as Kant's suggestion that human history is teleologically ordered towards increasing civilization, cultivation, and moralization (7:342)).

Foucault (like Nietzsche but unlike Kuhn) explains the emergence, modification, and reinterpretation of various epistemic fields in terms of power or domination. While Kuhnian paradigms are relatively benign structures of thought, and Kantian categories are necessary conditions that function to empower humans to know the world, Foucaultian epistemes are moves in a "hazardous play of dominations" (Foucault 1984: 83, cf. Foucault 1977: 27). The simplistic way of understanding Foucault, as merely pointing out that people often advance agendas by trying to get others to think like them, fails to get Foucault's more complex point. For one thing, precisely because it is so ubiquitous, Foucault does not see power as intrinsically problematic (See Foucault 1982:220). Foucault's approach to power does not center on its use by some human agents to dominate and control other human agents. For Foucault, power involves systems of knowledge and action that constrain and enable further knowing and acting (Foucault 1977: 27-8, see too Foucault 1984:150). Whereas Kuhn emphasizes the role that individual scientists play in the emergence of new scientific paradigms, Foucault focuses on larger social and institutional forces that are both "made up" of human actions and also "determine the forms and possible domains" of human thought and action. Thus, for example, "a certain way of rendering men docile and useful . . . required the involvement of definite relations of knowledge . . . [and thereby] . . . made the human sciences historically possible" (Foucault 1977: 305). The power structures of

²³⁴ Foucault scholars often distinguish between a period within which archeology is Foucault's primary method of history (beginning with *History of Madness* and ending with *The Order of the Things* and *The Archeology of Knowledge*) and a period within which genealogy dominates his approach (in *Discipline and Punish* and *The History of Sexuality*). But although Foucault does not explicitly articulate genealogy as governing his approach until "Nietzsche, Genealogy, History" (1971) elements of both methods are evident through his work. The shift between early and later works is primarily one of emphasis.

modern society shifted from an emphasis on establishing the absolute power of the king over subjects to punishing free and equal citizens to establishing “normality” in a population. The present “carceral society” that seeks “docile and useful” human bodies depends upon knowledge of human beings as subject-objects capable of (self-)control/responsibility and allows for techniques of observation that make the construction of this sort of knowledge possible. This does not mean that “human sciences emerged from the prison” (Foucault 1977: 305); Foucault is not claiming some sort of plot on the part of political leaders to set up empirical human sciences in order to better control their subjects. His point is not about some particular human agents dominating over others, but about a power-knowledge *system* that requires both certain forms of domination and certain forms of knowledge (cf. Foucault 1990:95, Foucault 1980: 203).

This de-centering of the subject as a locus of power and knowledge arises from Foucault’s historicizing of the very notion of a human “subject.” For Foucault, the subject itself is a recent historical emergence, a part of our present episteme, and one the contingency of which Foucault aims to reveal. Foucault’s work as a whole is a “history of the different modes by which, in our culture, human beings are made subjects” (Foucault 1982: 208) While revealing the contingency of this conception of man as subject-object, Foucault refuses simply to adopt the episteme he analyzes. Instead, his genealogical methodology effects a new way of studying power and knowledge, one that does not depend upon “man” as subject.

As a result, Foucault’s historical a priori is much more radical than Kuhn’s. Kuhnian paradigms can be understood as historically shifting a priori structures of cognition and one might even historicize Kant’s transcendental anthropology to take them into account. But Foucault calls into question the whole transcendental perspective, the whole idea of the thinking subject as the locus of cognition/knowledge.

[Genealogy] needed to be something more than the simple relativization of the . . . [transcendental] ²³⁵ subject. I don’t believe the problem can be solved by historicizing the subject . . . , fabricating a subject that evolves through the course of history. One has to dispense with a constituent subject, to get rid of the subject itself, that’s to say, to arrive at an analysis which can account for the

²³⁵ Foucault says “phenomenological” here since his targets are Husserl and Heidegger, but he says virtually the same thing about Kantian transcendental subjects. See, e.g., Foucault 1969/1982: 128 for a similar claim about the “formal a priori.”

constitution of the subject within a historical framework. And this is what I would call genealogy, that is, a form of history which can account for the constitution of knowledge, discourses, domains of objects, etc., without having to make reference to a subject which is either transcendental in relation to the field of events or runs in its empty sameness throughout the course of history. (Foucault 1984: 58-9, see too 73)

Foucault's historical a priori is not a subjective from-within perspective that changes through different historical conditions. For Foucault, neither knowledge nor the a priori are primarily "within" subjects. Knowledge is part of a knowledge-power structure that constrains and includes human actions; the "subject" (of knowledge) that is the focus of Kant's transcendental anthropology (of cognition) is, for Foucault, a recent innovation of our present knowledge-power complex, an innovation wrapped up with domination in the service of docile normalcy, an innovation that – in theory but especially in historical practice – Foucault seeks to resist and reinterpret. By describing how "we constitute ourselves as subjects of knowledge" (Foucault 1983: 237), Foucault de-privileges Kant's "transcendental" perspective, reinterpreting it as a contingent perspective created by historically local power relations (cf. Foucault 1973: 310, 322).

While the human "subject" is one way of describing the notion of "man" at the heart of Foucault's project, this "subject" is not merely the subject of Kant's transcendental anthropology. Foucault has in mind the whole "empirico-transcendental doublet" that seemingly characterizes Kant's anthropology as a whole. For Foucault, Kant inaugurated a shift from the Classical conception of thinking as essentially representation with a view of human cognition as something that orders the world in terms of its *own* nature. But for Foucault, this shift leads to a problem, since "man appears in his ambiguous position as an object of [empirical] knowledge and as a subject which knows" which leads, at least in its 19th and 20th century forms, to an "analytic of finitude," where thinkers aim to show how "man's being will be able to provide a foundation in their own positivity for all those forms that indicate to him that he is not infinite" (Foucault 1973:312, 315). In the end, Foucault argues that this analytic is irresolvable, that the 19th and 20th centuries represent a series of failed attempts to analyze the human being as "a being such that knowledge will be attained in him of what makes all knowledge possible" (Foucault 1973:318, see too 322). The result an intellectual culture that "produces, surreptitiously and in advance, the confusion of the empirical and the transcendental . . . [a]nd so we find philosophy falling asleep once more . . . , this time not in the sleep of Dogmatism but that of

Anthropology [where a]ll empirical knowledge, provided it concerns man, can serve as a possible philosophical field in which the foundation of knowledge, the definition of its limits, and, in the end, the truth of truth must be discovered” (Foucault 1973:341). Foucault thus suggests a shift away from “man” as object of anthropological investigation towards a “Nietzschean . . . promise of the superman,” a refusal – “with a philosophical laugh” – to give into the myth of “man” (Foucault 1973:342-3).²³⁶

While Kuhn’s historical analyses of sciences primarily call into question Kant’s transcendental anthropology of cognition, Foucault’s history of “man” challenges virtually every aspect of Kant’s anthropology. Archeological and genealogical investigations of how empirical human sciences are caught up in systems of domination and control reveal the contingency and potential dangers of these sciences. These analyses challenge naturalist attempts to use empirical study of human beings as comprehensive answers to the question “What is the human being?”, but they also call into question Kant’s own empirical (and pragmatic) anthropology, which, like its more contemporary forms, depends upon classification and observation in the service of normalization and control. Moreover, Foucault’s genealogical treatments of sexuality – which quickly evolve into a wholesale genealogy explaining how “we constitute ourselves as moral agents” (Foucault 1983: 237) – aim to show that the way we (and Kant) think of ethics is historically local. On Foucault’s account, “[N]obody is obliged in classical ethics . . . But, if they want to have a beautiful existence, . . . a good reputation, . . . to be able to rule others, they have to do” certain things (Foucault 1983: 240). Kant’s transcendental anthropology of volition starts from the “fact” that human beings are morally obligated and reasons from that fact to the basic structure of obligation (the categorical imperative) and the conditions of possibility of responsibility (transcendental freedom). If moral responsibility itself is merely an aspect of modern European knowledge-power, Kant’s moral philosophy merely answers the question “What ought I – as an 18th century modern man – do?” and his “anthropology” is really just a study of human beings living within a particular,

²³⁶ Strikingly, however, Foucault recognizes that this criticism of “anthropology” and the “empirico-transcendental doublet” does *not* apply to Kant himself. Not only is “surreptitious . . . confusion of the empirical and the transcendental” enacted “even though Kant had demonstrated the division between them” (Foucault 1973:341), but Foucault rightly notes that “The Kantian moment,” rather than being a part of the Modern Age and susceptible to its problems, is rather “the link between the two [the Classical and the Modern ages]” (Foucault 1973:343). Moreover, Foucault’s own project, as much as he might want to escape the modern episteme of empirico-transcendental doublet, falls prey to precisely the confusion he notes in other historicist thinkers, that of using empirical investigation of human beings – in Foucault’s case, of historical epistemes – to make transcendental claims about human norms and possibilities.

contingent system of knowledge-power.²³⁷ The moral law is “universal” only in the sense that it represents a particularly modern-European ambition to subordinate all diversity and particularity (“savagery” and “deviance”) into a single overarching system of normalcy and control. And “autonomy” is really just a way in which the observational systems of the modern world seek to impose power through creating human beings who regulate themselves. For Foucault, “Kant introduces one more way in our tradition whereby the self is not merely given but is constituted in relationship to itself as subject” (Foucault 1983: 252)

In the end, even the very question “What is the human being?” is suspect: “the notion of human nature seems to me mainly to have played the role of ...designating certain types of discourse....” (Foucault 1984: 4). “[M]an, as a primary reality with its own density, as the difficult object and sovereign subject of all possible knowledge” is a recent innovation, something that “has no place” even in the Classical era of Descartes, much less in ancient or medieval forms of human life (Foucault 1973: 310). And even if the question could make sense, Kant’s answer to it is, at best, the careful analysis of a particular episteme that arose during the 18th century and still plays a substantial role in our self-conceptions, the human being as “empirico-transcendental doublet.” Moreover, the purpose of Foucault’s “genealogy of the subject” is not, as in Kant’s “Conjectures on the Beginning of Human History,” to show the emergence of a from-within perspective that is now, for all intents and purposes, the necessary structure of human beings as such. Instead, Foucault aims to disclose the contingency of human subjectivity as it emerged in our culture in order to open up possibilities for revising who we are:

Maybe the target now is not to discover what we are, but to refuse what we are . . . We have to promote new forms of subjectivity through refusal of [the] kind of individuality which has been imposed on us for several centuries. (Foucault 1982: 216)

Rather than transcendental justification of the structure of subjectivity through analysis of its conditions of possibility, Foucault offers a genealogy of the emergence of our distinctive forms of subjectivity in order to refuse those forms. Summing up his relationship with Kant, Foucault explains,

²³⁷ One might even wonder whether *Kant’s* answer to “What is the human being?” can continue to be *our* answer. If Kant’s conception of the subject is as far from the Greeks’ as Foucault suggests, it might reasonably be seen as equally far from our own conceptions. Insofar as this is true (and it is certainly true to some extent), Foucault deserves no small part of the credit/blame for our own transformed understandings of subjectivity, freedom, and responsibility.

If the Kantian question was that of knowing what limits knowledge has to renounce transgressing, it seems to me that the critical question today has to be turned back into a positive one: in what is given to us as universal, necessary, obligatory, what place is occupied by whatever is singular, contingent, and the product of arbitrary constraints? The point is, in brief, to transform the critique conducted in the form of necessary limitation into a practical critique that takes the form of a possible transgression.

This entails and obvious consequence: that criticism is no longer going to be practiced in the search for formal structures with universal value, but rather as a historical investigation into the events that have led us to constitute ourselves and to recognize ourselves as subjects of what we are doing, thinking, saying. In that sense, this criticism is not transcendental, and its goal is not that of making a metaphysics possible: it is genealogical in its design and archaeological in its method.

Archaeological – and not transcendental – in the sense that it will not seek to identify the universal structures of all knowledge or of all possible moral action, but will seek to treat the instances of discourse that articulate what we think, say, and do as so many historical events. And this critique will be genealogical in the sense that it will not deduce from the form of what we are what it is impossible for us to do and to know, but it will separate out, from the contingency that has made us what we are, the possibility of no longer being, doing, or thinking what we are, do, or think. It is not seeking to make possible a metaphysics that has finally become a science; it is seeking to give new impetus, as far and wide as possible, to the undefined work of freedom. (Foucault 1984: 45-6)

III. Human Diversity Today

Historicism is one important form of a general contemporary trend towards emphasizing human diversity. For historicists, there can be no uniform answer to the question “What is the human being?” because human beings change throughout time. But even at any given time, humanity includes substantial diversity. The most important kinds of diversity to receive attention over the past 200 years are differences based on sex or gender, race or ethnicity, and culture. In chapter five, we discussed Kant’s (over)attentiveness to such distinctions and looked at some contemporary responses to Kant’s accounts of sex, gender, race,

and ethnicity. This section focuses on the important role that increased awareness of cultural difference plays in conceptions of “the” human being. Recognizing and appreciating the distinctness of various different human groups is not a wholly new phenomenon. Kant was writing after over 200 years of European exploration of the globe, and he was a voracious reader of the travel literature of his day. Montaigne, one of Kant’s literary heroes, wrote essays defending the practices of “cannibals” in the Americas, exposing the culture-boundedness of European prejudices, and promoting greater tolerance and understanding of human diversity. But with increased globalization over the past hundred years, diversity has become more obvious, more important, and more endangered than ever before.

Increased awareness of human diversity and interactions between groups with different cultures brings with it both practical and philosophical challenges. Like historicism, the awareness of human diversity threatens Kant’s anthropology with relativism. If some cultures do not ascribe moral responsibility to one another, or do not see morality as universal, that would call into question Kant’s moral philosophy. If some cultures do not understand the world in terms of necessary causal interactions between spatial-temporal objects, that would threaten his epistemology. Kant’s empirical anthropology, with its tidy classification of human mental states, is based upon the introspection and limited observations of a man who never went more than 90 miles from home. Today one might wonder whether there really are empirically universal characteristics of human nature. Even Kant’s conception of historical progress, which sees human “unsocial sociability” driving human progress, might seem incompatible with the existence of relatively peaceful and stable cultures.

Practically, the problems are equally severe. What responsibilities do we have with respect to those from other cultures? For Kant, enlightenment arises through the vibrant exchange of ideas aiming towards the truth. But this truth-orientation is also an orientation towards finding agreement, or conformity. As “multiculturalism” and “pluralism” have come to be taken as goods in their own right, one might question Kant’s emphasis on reaching agreement. Might it not be better to allow, protect, and even promote widely divergent conceptions of reality amongst different human groups. Efforts to preserve dying languages can be seen as part of this general interest. Morally, the problems are even more severe. If “we” share, broadly speaking, Kant’s commitment to respect for humanity, how should this respect manifest itself in interactions with those from other cultures? If women in a particular culture are mistreated or abused, should “we” refrain from

interfering out of respect for the culture or should we intervene out of respect for the woman? Or is there some other option? Politically, the problems of diversity are acute. Kant argued that no political order is fully just until it becomes part of a global federation of states. Since our planet is finite and interconnected, no set of people has an absolute right to assert sovereignty over its territory without getting some sort of consent from all other people on earth. Kant used this insistence upon consent to argue against unjust colonization of other peoples that was widespread during the 18th century.²³⁸ But he also recognizes that his own theory of universal consent is susceptible to an all-too-common “Jesuitism” that would “ask whether . . . we should not be authorized to found colonies, by force if need be, in order to establish a civil union . . . and bring these human beings (savages) into a rightful condition” (6:266).

One popular way of reacting to human diversity is to embrace cultural relativism about truth, virtue, and beauty. As in the case of historicism, there are both methodological and substantive versions of this relativism. Methodologically, at least a limited relativism has become the norm among contemporary cultural anthropologists. In studying other cultures, anthropologists typically focus on discerning the practices, presuppositions, and values of a particular culture without aiming to assess the value of those practices in terms of any supposed absolute standard. Clifford Geertz, arguably the preeminent contemporary cultural anthropologist, makes explicit a “relativist bent” that is “in some sense implicit in the field as such” (Geertz 2007:44). And Ruth Benedict, in her classic *Patterns of Culture* (Benedict 1934/2005), explains:

To the anthropologist, our customs and those of a New Guinea tribe are two possible social schemes for dealing with a common problem, and insofar as he remains an anthropologist he is bound to avoid any weighing of one in favor of the other. (Benedict 1934/2005: 1)

Methodologically, the relativist bent of anthropologists commits them to engaging in a different project than that of philosophers or even many psychologists. Rather than trying to figure out the best way of dealing with various problems that might arise in societies (whether one’s own or that of another), the anthropologist observes and seeks to understand how other cultures see and respond to those problems. To avoid projecting one’s values onto other cultures and to remain sensitive to the nuances of other cultures, some sort of relativism – at least in the

²³⁸ (See PP 8:357-60, MM 6:353)

negative sense of refraining as much as possible from evaluating other cultures in terms of one's own – has proven immensely valuable for understanding human diversity.

Substantive relativism takes this relativist bent further, claiming that basic concepts of truth and value are culture-relative not merely in the sense that what people *find* true and valuable is largely culture-bound, but that truth and value are *in fact* culturally relative. Methodological relativism simply refrains from asking ultimate questions about Truth or Goodness. Substantive relativism claims that there are no (universal) answers to those questions, that “morality differs in every society, and is a convenient term for socially approved habits,” so “we do not any longer make the mistake of deriving . . . morality . . . directly from . . . human nature. We do not elevate it to the dignity of a first principle. (Benedict 1934) Geertz too sees relativism not merely as a methodological presupposition of anthropological research but as a necessary *consequence* of anthropological insight:

One cannot read too long about Nayar matriliney, Aztec sacrifice, the Hopi verb, or the convolutions of the hominid transition and not begin at least to consider the possibility that, to quote Montaigne . . . , “each man calls barbarism whatever is not his own practice . . . for we have no other criterion of reason than the example and idea of the opinions and customs of the country we live in.” That notion, whatever its problems, and however more delicately expressed, is not likely to go away unless anthropology does. (Geertz 2001: 44-5)

As much as Geertz and Benedict see substantive relativism as a consequence of anthropological study, however, it does not follow logically from the fact of human diversity, nor from the methodological relativism that enhances our understanding of that diversity. It would be perfectly sensible to recognize that human beings hold different beliefs about physics, for example, and even to study different cultures' mathematical systems without evaluating their soundness, while still affirming that, for instance, projectiles really do travel on parabolic paths and spells cannot transform lead into gold. Similarly, it would be perfectly sensible to recognize that human beings have different sets of moral values regulating interactions, while still affirming that, for example, value systems within which malicious deception is condoned or slavery is promoted are wrong.

Moreover, in its most blatant form, substantive relativism is simply self-defeating. Benedict rightly notes that “the recognition of cultural relativity carries with it its own values,” and while these “need not be

those of the [prior] absolutist philosophies” (Benedict 1934/2005: 278), they are just as absolutist as those.²³⁹ As anthropologist D. Sperber has noted, “while the best evidence for relativism [is] in the writings of anthropologists, the best evidence against relativism is . . . the very activity of anthropologists” (Sperber 1982: 180). Strikingly, substantive relativism turns out to be a particularly parochial value system; while there may be *some* other cultures that believe in epistemic and moral relativism, most do not, and the particular forms of relativism dominant today have emerged only in the context of moral and anthropological developments in “the West” in the past hundred years. Substantive relativism of this dogmatic sort is just as ethnocentric and absolutist as any other substantive dogma.

There are, however, less dogmatic sorts of relativism that still go beyond the merely methodological relativism of contemporary anthropology. For one thing, anthropological study, like the historical studies of Foucault and Kuhn, can show that the from-within perspective that Kant took for granted is not as universal as he supposed. Geertz makes this connection explicit. Citing his own pioneering work in cultural anthropology, he notes,

[T]he constructivism of Thomas Kuhn and . . . Foucault . . . suddenly made a concern with meaning-making an acceptable pre-occupation for a scholar to have [and] they provided the . . . speculative instruments to make the existence of someone who saw human beings as, quoting myself . . . “suspended in webs of meaning they themselves have spun” a good deal easier . . . [In] Bali, . . . I tried to show that kinship, village form, the traditional state, . . . and, most infamously, the cockfight could be read as texts, or to quiet the literal-minded, “text-analogues” – enacted statements of . . . particular ways of being in the world. (Geertz 2001: 17)

When we look back at Geertz’s anthropological work, the connections become clear. Even in 1973, Geertz emphasized that “The whole point of [his] . . . approach to culture is . . . to aid us in gaining access to the conceptual world in which our subjects live so that we can, in some extended sense of the term, converse with them” (Geertz 1973: 24). In that context, “descriptions of Berber, Jewish, or French culture must be cast in terms of the constructions we imagine Berbers, Jews, or Frenchman to place upon what they live through” (Geertz 1973: 15). In such a context, one comes to see our own “ideas, our values, our acts,

²³⁹ Benedict herself often appeals to her own “objective point of view” (e.g. Benedict 1934: 7-8) on other cultures, drawing on a non-relativist epistemic ideal in defense of her anthropological method.

even our emotions [as] cultural products” (Geertz 1973: 50). Putting ourselves in other “worlds” makes Kant’s kind of transcendental anthropology impossible, not because of a dogmatic assertion that there are no universals, but simply because we come to see new possibilities – and hence new “conditions of possibility.” One need not deny a universal point of view to make one’s own point of view seem provincial. And a transcendental analysis of a provincial and contingent point of view falls short of the anthropology in which Kant was so interested.

As in the case of historicism, of course, this very *general* sort of objection is most important when fleshed out in terms of *specific* points of comparison between what Kant took to be universal aspects of human nature and what anthropologists find in other cultures. For example, Ruth Benedict – drawing on the work of Reo Fortune – describes the Dobu people of the South Pacific in ways particularly problematic for Kant’s moral philosophy:

The Dobuan is dour, prudish, and passionate, consumed with jealousy and suspicion and resentment. Every moment of prosperity he conceives himself to have wrung from a malicious world by a conflict in which he has worsted an opponent. The good man is the one who has many such conflicts to his credit, as anyone can see from the fact that he has survived with a measure of prosperity. It is taken for granted that he has thieved, killed children and his close associates by sorcery, cheated whenever he dared. [T]heft and adultery are the object of the valued charms of the valued men of the community. (Benedict 1934/2005:168-9)

As troubling as is the general fact that moral norms might differ from one context to another, the *details* of these differences seem to pose particular problems for Kant. Whatever the Dobuans mean by the “good” man, they cannot mean the man who acts only on maxims that he can will to be universal. A “good” man, on Benedict’s reconstruction of the Dobu people, is one who acts on maxims that precisely *cannot* be universalized, one who exploits and misuses his fellows for his own benefit. And thus any “transcendental anthropology of volition” for the Dobu will, it seems, have to look very different from Kant’s.

Along with this undoing of the foundations of Kant’s universalist anthropology, many of those interested in human diversity add an ethical and pragmatic “relativist bent.” Geertz, for example, largely accepts the point that substantive relativism is self-undermining. But he takes this precisely as a reason not to worry about objecting to it. As he puts it, “The image of vast numbers of anthropology readers running around in

so cosmopolitan a frame of mind as to have no views as to what is and isn't true, or good, or beautiful, seems to me largely a fantasy" (Geertz 2001: 46). By contrast, however, the thought of lots of Kantians running around interpreting everyone in terms of their own prejudices about knowledge, goodness, and even beauty is one worthy of genuine concern: "provincialism . . . [is a] more real concern [than relativism]" (Geertz 2001: 46). And in that context, studying different ways of being in the world facilitates a sort of humility about one's own ways, an openness to possibilities, a valuing of differences. Especially in a world where "alien turns of mind are mostly not really elsewhere, but alternatives for us, hard nearby" (Geertz 2001: 83), we do not need a Kantian, *universalist* anthropology but rather a proliferation of "connoisseur[s] . . . of alien turns of mind," of whom "the connoisseur par excellent . . . has been the ethnographer (the historian too, to a degree, and in a different way the novelist, but I want to get back on my own reservation), dramatizing oddness, extolling diversity, and breathing broadmindedness" (Geertz 2001: 82-3).

The undermining of the possibility of Kantian naïveté about the universality of one's own perspective, combined with an ethical sense of the importance of broadmindedness, tolerance, and understanding others, leads to a new vision for how to go about answering the question, "What is the human being?"

If we want to discover what the human being amounts to, we can only find it in what human beings are: and what human beings are, above all other things, is various. It is in understanding that variousness – its range, its nature, its basis, and its implications – that we shall come to construct a concept of human nature that, more than a statistical shadow and less than a primitive dream, has both substance and truth . . . To be human here is thus not to be Everyman; it is to be a particular kind of man, and of course men differ . . . [I]t is in a systematic review and analysis of [different ways of being human] – of the Plain's Indian's bravura, the Hindu's obsessiveness, the Frenchman's rationalism, the Berber's anarchism, the American's optimism – that we shall find out what it is, or can be, to be a man. (Geertz 1973: 52-3)²⁴⁰

²⁴⁰ Throughout this quotation, I've changed "man" and "men" to "human being" or "human beings." Importantly, too, Geertz adds that this "series of tags" is one that "I should not like to have to defend" and, further, that "we must . . . descend in detail, past the misleading tags, past the metaphysical types, past the empty similarities to grasp firmly the essential character of not only the various cultures but the various sorts of individuals within each culture, if we wish to encounter humanity face to face" (Geertz 1973: 53).

Even without going as far as substantive relativism, contemporary cultural anthropology – and its spin-offs into cultural studies, postcolonial studies, gender studies, and studies of diversity in all its forms – invites a shift in perspective. Just as historicists argue against a single notion of “human being” for all times, cultural anthropologists object to a cross-cultural conception of *the* human being, replacing it with rich variety. And as in the case of historicism, while there are potential implications of cultural anthropology for his empirical anthropology, the cultural-anthropological threat to Kant is most acute in the context of his transcendental anthropology.²⁴¹

IV. Kantian Responses to Historicism

Kuhn, Foucault, Benedict, and Geertz all draw attention to fundamental human differences that not only reflect empirical variations but also affect how humans see and live in their worlds. They threaten not only Kant’s empirical claims, but his transcendental anthropology. While the specific claims of each thinker would require very specific responses, there are three general strategies that a contemporary Kantian might use to respond to these challenges, what I will call “sticking to one’s guns,” “strategic retreat,” and “surrender.”

Sticking to one’s guns

The “guns” to which Kant would stick in these cases are the basic tenets of his transcendental anthropology. A gun-sticking Kantian would refuse to give up the central claims that all human volition involves an awareness of the moral law and all human cognition involves spatial-temporal intuition and a priori cognitive categories such as causation. With respect to challenges posed by scientific developments, this would involve denying that relativity theory and quantum mechanics, as generally interpreted, provide actual cognition of the world. Any contradiction between Kant’s transcendental anthropology and Einstein’s physics would be bad news for Einstein, not for Kant. With respect to Foucault and Geertz, Kant might simply deny that these figures properly interpreted human history or diverse cultures, or he might deny that the

²⁴¹ Empirically, cultural anthropologists do not challenge as much of Kant’s conception of human beings as one might imagine. Virtually none doubt that people in other cultures have cognitions, desires, or capacities to develop habits. And much recent work shows many universal human characteristics (see Pinker 2002 and Spiro 1987). The real challenge comes from the way in which cultural diversity can undermine Kant’s from-within accounts of thought, feeling, and action.

interpretations really represent counter-examples to his transcendental anthropology.

While sticking to one's guns might seem simply pig-headed, it is not wholly unjustified. Regarding developments in science, Kant's arguments for the aprioricity of our forms of cognition were based on conditions for arriving at genuine empirical *understandings* of the world. And it is not entirely clear that modern scientific theories are literally understandable in their non-Kantian forms. The mathematics of relativity theory and quantum mechanics does not conflict with Kant's a priori structures of human cognition. Kant never claimed that it would be impossible to *think about* what would follow from rejecting one or more of Euclid's axioms, only that it would be impossible to actually cognize such a world, that is, to fill in one's concepts with intuitions of objects. And when "explaining" objects in quantum-mechanical or relativist terms, scientists notoriously turn to metaphor and analogy, which suggests that a literal understanding of modern physics is not yet forthcoming. For a gun-sticking Kantian, intuitive comprehensibility would be a constraint on scientific realism, and Kant himself proposes a similar approach against those in his day (and ours!) who assume the existence of a vacuum.

Nearly all natural philosophers, since they perceive a great difference in the quantity of matter of different sorts in the same volume . . . infer that this volume . . . must be empty in all matter, although to be sure in different amounts. But . . . their inference rest[s] solely on a metaphysical presupposition . . . for they assume that the *real* in space . . . is *everywhere one and the same* and can be differentiated only according to its . . . amount. Against this presupposition, . . . I oppose a transcendental proof, which, to be sure, will not explain the variation in the filling of space, but which still will entirely obviate the alleged necessity of the presupposition . . . which has the merit of at least granting the understanding the freedom to think of this difference in another way. (A173-4/B215-6)

It is unnecessary to enter into the details of this particular debate here; the general point is that Kant is willing to set his transcendental proof against the "unanimous" decrees of "nearly all natural philosophers," and to do so even when he cannot himself explain the phenomena that their theories purport to explain. Similarly, a Kantian today might insist that quantum mechanics and relativity theory cannot be adequate explanations of the world, since they are inconsistent with our forms of intuition and thus literally incomprehensible as applied to objects. They can still be good models for prediction, but not for understanding, and

rejecting scientific realism here may even “have the merit” of encouraging work in new directions in physics. Akin to Einstein’s early critique of quantum mechanics, we might see Kant as insisting that the fact that scientists have not *found* deterministic laws does not imply that there *are* no deterministic laws.²⁴² Kant might even go further, pointing out that appeals to sub-atomic quantum states and bendable space-time that cannot be literally understood in the way we understand objects of experience reflects a 21st century version of the classic metaphysical temptation to turn to things-in-themselves – thinkable but non-intuitable psuedo-objects – as a shortcut for explaining the empirical world.²⁴³

Similarly, with respect to Foucaultian genealogy and Geertzian anthropology, sticking to his guns is more plausible than it might at first appear. One approach here is an attitude towards cultural difference that many think Kant actually adopted. One might maintain Kant’s transcendental anthropology while still admitting human diversity by simply insisting that some peoples and cultures are not “human” in the fullest sense.²⁴⁴ If a particular group conceives of decision-making *purely* in categories of beauty (as Foucault suggests for the ancient Greeks) or prudence (as Benedict suggests for the people of Dobu), then Kant might just say that such people lack a fully developed predisposition to personality, lack a genuine moral sense, and, in that sense, are not really “human.” Whether or not this would warrant treating them with disrespect would remain an open question, but it would be a way of saving Kant’s *philosophical* account from anthropological challenge. As offensive as the approach sounds, there is some degree to which it is unavoidable. We do see the world through our own eyes and our own values, and while our perspective might change through understanding others, there simply is not – if relativist cultural anthropology is right – a single perspective that all groups share. But since we do need to decide what to believe and how to act, we will, at least in practice, think and act in accordance with norms we think best, and we will thereby at least implicitly view other groups as seeing through the wrong eyes. Even the relativist bent shared by Benedict, Geertz and Foucault is a particular bent developed in a particular culture, one not shared by many other

²⁴² See Einstein et. al. 1935, but cf. Bohm 1951 and Bell 1964, 1987. For discussion, see Baggott 2004 (which explains how experimental work has largely vindicated indeterminist interpretations of quantum mechanics against the Bohm-Bell deterministic interpretation) and Cushing 1994 (which offers a sustained defense of Bohm’s deterministic interpretation).

²⁴³ For a Kantian argument that treats unobservables in science in roughly this way, see Hanna 2006.

²⁴⁴ For a treatment of Kant’s accounts of diversity as a form of “alienology,” see Cohen 2010.

people in many other cultures. Insofar “What is the human being?” is a *normative* question about how best to be human, it is almost inevitable, if we know enough about human diversity, to see at least some forms of diversity as failures to live up what it means to be a human in the fullest sense.

But there are other, less judgmental, ways for a Kantian to stick to her guns. One important move for the Kantian will be to distinguish between *particular* knowledge or value claims and the overall *structure* of knowledge- or value-claims. Even if other cultures, for example, disagree about what causes particular kinds of changes, they may still agree on the notion of temporal succession (and thereby on some shared conception of causation). Even if cultures disagree about, say, cannibalism, they might still adhere to some general conception of respect for others. And even if – as in Benedict’s account of the Dobu – some cultures do not even believe in anything that could be called Kantian respect, they might still adhere to a sense of “goodness” as something that would be “good” for anyone. Thus one of Benedict’s Dobuans might say that *anyone* who successfully exploits and abuses others is “good.” And in such a case, Kant might be able to run transcendental arguments to show that built into this conception of goodness is a standard of goodness at variance with the particular ethical prescriptions of Benedict’s Dobu.

Kant also could rightly insist upon a difference between from-within standpoints of evaluation and deliberation and a people’s actual customs and practices. Kant admits that human beings are “radically evil” and even that this evil manifests itself in corrupting social networks to the point that “someone already counts as good when his evil is common to a class” (6:33). Moreover, given variable conditions, human inclinations and prejudices develop in different ways, with correspondingly different manifestations of immorality in different cultures. Thus just as the Dobu might excuse the immorality of adultery and witchcraft and praise its prudence, those on Wall Street might excuse the immorality of competitive “sharp practice” and praise the returns one brings home to one’s shareholders. In neither case are these forms of praise reflections of a different moral code; they just reflect the ordinary way in which social forms of radical evil corrupt our strict application of the moral law we recognize from-within.

Finally, Kant might rightly point out that much of the perceived variation amongst cultures could be due to a *prejudgment* or inclination towards novelty, one widely shared by the sorts of people that typically become anthropologists and (Foucauldian) historians. Kant emphasizes,

from a multiplicity of descriptions of countries one can prove, if one wants to, that Americans, Tibetans, and other genuine Mongolian peoples have no beard, but also, if it suits you better, that all of them are by nature bearded . . . ; that Americans and Negroes are each a race, sunk beneath the remaining of the human species in their mental predispositions, but on the other side by just as apparent records that as regards their natural predispositions, they are to be estimated equal to every other inhabitant of the world; so it remains to the choice of the philosopher whether he wants to assume differences in nature or wants to just everything in accordance with the principle “Everything is as it is with us.” (8:62)

As we saw above, a certain relativistic bent – and often even an orientaling fascination with the exotic (see Said 1979, Obeyesekere 1993, 2005) – is not only a natural *result* of anthropological study, but a sort of dispositional and methodological *presupposition* of certain kinds of anthropological investigation (including, alas, Kant’s). Likewise, Foucault’s genealogical and archaeological projects, as much as they seem to provide evidence of the historical emergence of frameworks of thought and action, ultimately *presuppose* a historicist approach to structures of human knowledge-power. In his own *Anthropology*, Kant points out that “without . . . a plan . . . all acquired knowledge [of the world] can yield nothing more than fragmentary groping around and no science,” and for Kant, this plan requires that “*General* knowledge always precede *local* knowledge,” that is, that one have a sense of the human being in general before studying local variations (7:120). Whether or not one agrees with *this* methodological prescription, it is worth noting that anthropologists studying the world with this sort of Kantian methodology may come to very different conclusions than those who *begin* with a more relativistic bent. There is reason for at least some skepticism about the empirical findings – and their interpretations – offered by the more relativistically-inclined anthropologists amongst us.

Thus whereas Foucault focuses on the *emergence* of contemporary notions of subjectivity, a Kantian historicist might instead look for Kantian conceptions of subjectivity in historical periods when Foucault denies them and in cultural contexts where contemporary anthropologists claim not to find them. Where Foucault claims that “nobody is obliged in classical ethics,” Kant might claim that classical notions of “beautiful existence” are taken, even in ancient Greece, as “to-

be-chosen” in ways that correspond to categorical “obligation.”²⁴⁵ Historical ways of describing subjectivity would be merely different “formulae” for common underlying transcendental structures of human beings that Kant elucidates in terms of “obligation,” “freedom,” and “autonomy.” It is worth emphasizing here that Kant’s moral anthropology and conception of subjectivity were at variance even with what one may have discovered – as an anthropologist or historian – about *his own time*. Kant emphasizes the independence of conscience from religion, the importance of adhering to strict principles, the extent to which morals must be carefully distinguished from the pursuit of happiness. Kant saw these claims as *implicit* in the volitional structure of his compatriots (and human beings in general), but all of them could have been occluded in a historical or anthropological study of his culture.²⁴⁶

This skepticism, of course, can lead to a different way of doing anthropology, but time must still tell whether this sort of Kantian cultural anthropology could hold up to the empirical facts on the ground. There are, however, promising hints that an anthropology that leaves more room for human universals – especially of the rational variety in which Kant would be most interested – may be more fruitful than anthropologists like Benedict and Geertz suggest.²⁴⁷ Perhaps the most famous example in recent anthropology is the debate between two of the preeminent anthropologists of our time – Gananath Obeyesekere and Marshall Sahlins – regarding the Hawaiians’ reception of Captain Cook. While Sahlins undertook that study largely with the “relativist bent” of a European interested in the exotic and developed a picture of native Hawaiians that makes them seem very different from Europeans then and now, Obeyesekere went into the study with a skepticism about the nature of European anthropological and “myth-making” practices and called into question exoticizing descriptions of the reception of Cook

²⁴⁵ Kant’s ethics, in fact, is often referred to as “deontological,” from the Greek term “*dei*,” which refers to that which is binding or what it behooves one to do, as in the Iliad’s “*ti de dei polemizemenai . . . Argeious*” (why do the *Argives* have to fight?) (II.9.337).

²⁴⁶ The fact that Kant sees his philosophical account of the unconditional nature of morality as unrecognized (and thus unreflected) in even the best moral philosophy of the seventeenth and eighteenth centuries implies that the absence of similarly worked out accounts of moral obligation in the ancients would come as no surprise. Moreover, Foucault’s historical claims are far from non-controversial, and there is substantial empirical support not only for universal and transhistorical moral norms but also for some basic conceptions of subjectivity and obligation that are universal and transhistorical.

²⁴⁷ See, e.g., Bok 2002, Brown 1991, Goodenough 1970, Pinker 1993, Schwarz and Bilsky 1990, and Spiro 1954.

shared by modern anthropologists and ultimately even Hawaiians themselves.²⁴⁸

To take a less famous and more pointed example, Susanne Kuehling has recently published a book-length study of the Dobu people in which she makes clear not only the profound limitations of the original field research on which Benedict relied (by an anthropologist – Fortune – who had access to a single Dobuan for a mere month’s time and wrote with “the imperial attitude of his time” (Kuehling 2005:14)) but also how “Benedict’s travesty” (Kuehling 2005:136) warped even that “vastly oversimplified” account to portray an “even darker” picture of the Dobu “as an extreme example of human moral possibilities” (Kuehling 2005:16), a “characterization . . . that bears no resemblance to [the] Dobu” experienced by Kuehling over the course of several years of intense fieldwork. In sharp contrast to the “inverted morality” portrayed by Benedict, Kuehling highlights the Dobu’s “egalitarian ethic” (Kuehling 2005:117) and emphasizes that among the Dobu an “ethics of respect, self-discipline, and generosity are the keys to appropriate behavior” (Kuehling 2005:69). Of course, Kuehling’s own account, which highlights the “ethics of exchange” as a framework for understanding the Dobu, does not show that the Dobu share the general presuppositions of Kant’s moral anthropology, but she goes into her study with a different agenda, with a different “general knowledge,” and thus Kant would be unsurprised that she neither looks for nor finds the sort of “predisposition to personality” in which a Kantian anthropologist would be particularly interested. What she clearly shows, however, is the importance of taking *any* anthropological “counter-examples” to Kantian transcendental anthropology with a grain of salt. And she thus justifies at least a provisional Kantian sticking-to-one’s-guns, an unwillingness to take as given the “observations” and “facts” of anthropologists who go into the field with deeply non- or even anti-Kantian presuppositional frameworks.

Strategic retreat

²⁴⁸ A full discussion of this very interesting debate would take us far beyond the scope of this book. For a clear overview with references to further sources, see Borofsky and Kane 1997.

Strategic retreat involves giving up specific a priori structures, such as Euclidean space, Newtonian deterministic causation, or specific formulations of the categorical imperative, but preserving more general a priori structures, roughly corresponding to Kant's. Just as contemporary psychology requires revising details of Kant's empirical anthropology but not his overall framework, historicist or anthropological studies might require revising merely details of Kant's transcendental anthropology. Alternative, one might strategically retreat by limiting the *scope* of Kant's a priori structures, insisting that they underlie ordinary experience but not scientific or moral theorizing.

As with sticking-to-one's-guns, strategic retreat is more plausible and more significant than it might first appear. With respect to relativity theory, for example, Henrik Lorentz had developed an alternative to Einsteinian special relativity theory that is both "empirically equivalent" to it and makes use of an "essentially classical spatio-temporal structure" (Friedman 2001:87). The probabilistic causation of quantum mechanics already fits with a slightly modified version of Kant's approach to causation. For Kant, the fundamental role of causation is to preserve the directionality of time. Kant assumed that the only causal principle by means of which one could order the world would be a principle according to which, given some cause, its effect necessarily followed according to a deterministic rule. Quantum mechanics conceives of causation as a principle according to which, given some initial state, its succeeding state is determined by a probabilistic rule. While giving up the necessity of the effect given the cause would be an important shift from Kant's category of causation, it would not require giving up the basic structure of succeeding states following from previous ones according to rules.

One might imagine retreating further, such that one ends up with increasingly thin a priori structures of cognition. Kuhn himself argues for some form of this approach, defending general epistemic virtues such as accuracy, consistency, breadth of scope, simplicity, and fruitfulness in generating new findings as general characteristics of any good scientific theory (Kuhn 1977). Similarly, Kant might argue that whatever particular structures human beings use to interpret their world, they make use of certain a priori principles to guide empirical cognition in general. This strategy could be extended to variations in conceptions of the world discovered by anthropologists. While other cultures may allow for witchcraft, cycles of time, or radically different approaches to understanding the world, one might still find common basic structures underlying all these approaches. Even something like Foucault's analysis of earlier forms of subjectivity might leave common structures – say, some general notion of normativity or basic distinction between a

from-within and an objective perspective – that are common to different, historically-local ways of conceptualizing these structures.

Beyond changing particular claims while keeping more general ones, an important sort of strategic retreat, especially in the scientific context, might be the concession of historicism with respect to scientific cognition while preserving Kant's transcendental anthropology for what one might call ordinary cognition.²⁴⁹ Even if scientists now think of the world as involving probabilistic causation and non-Euclidean space, ordinary human experience is universally and ahistorically based on Kant's a priori cognitive structures. Especially conjoined with a deprivileging of scientific cognition, such a retreat preserves a substantial role for Kant's transcendental anthropology of cognition while leaving science subject to historicist interpretation. (One might do something similar with other cultures, arguing that, say, religious or mythical beliefs might violate Kant's categories of experience while everyday interactions would still be governed by them.) This sort of retreat is not Kant's, in that Kant saw his transcendental anthropology as providing metaphysical conditions of possibility not only of "experience" but also of "natural science."²⁵⁰ And this approach is not without its problems. For one thing, even ordinary human cognition seems less universal as Kant suggests, in part because ordinary experience is – like science – shaped by historically contingent cultural paradigms. Moreover, while Kant allows for distinctions between apparently contradictory perspectives (such as the empirical and the practical), he does so only after carefully explaining why those perspectives do not actually conflict with one another, why they can be taken to refer to different worlds, and how those different worlds relate to one another. To allow for legitimate scientific theories that seem to conflict with conditions of possibility of experience, Kant would have to show that this conflict is merely apparent and explain how the scientific understanding of the world relates to the commonsensical one. Such a project may be possible, but it represents a considerable burden for this sort of strategic retreat.

Surrender.

²⁴⁹ One might also, especially vis a vis something like Foucault's historicism, strategically retreat by limiting the scope of one's transcendental analysis to the present (and recent past), such that whatever morality looked like "from-within" for ancient Greeks, it involves moral responsibility and personal accountability *now*.

²⁵⁰ See especially his *Metaphysical Foundations of Natural Science*.

Strategic retreat might go so far that virtually nothing worth saving is left of Kant's transcendental anthropology. Even the most basic cognitive and volitional values might be exposed as historically contingent. In this context, one might simply need to concede that Kant's transcendental anthropology must be replaced by a historicist one. Even outright surrender, however, need not involve a wholesale rejection of the insights of Kant's transcendental anthropology. For example, arguably the most exciting Kantian philosopher of science today – Michael Friedman – has defended a “modified version of a Kantian philosophy of science” centered around the concept of a “relativized yet still constitutive a priori” (Friedman 2001:71). Friedman embraces the historical contingency of cognitive structures by which human beings make sense of the world while sustaining a commitment to the status of these structures as constitutive of (our understanding of) that world. For Friedman, even a relativized a priori is an important contribution to contemporary philosophy of science in that it implies, against Quine's holistic “fabric” of knowledge, that knowledge has a structure with “fundamental asymmetries,” such that within any (scientific) body of knowledge, there are “necessary presuppositions constituting the conditions of possibility of the properly empirical parts” (Friedman 2001: 35, 37). Even if every aspect of human knowledge is in principle revisable in the light of further experience, there is a fundamental distinction between the ways that specific empirical laws and the (relativized) a priori structures of cognition are revisable in the light of experience. Even when historicized, the general approach of Kant's transcendental anthropology contributes to understanding human beings. In fact, one might even argue, it is Kant's transcendental anthropology of cognition, and especially his emphasis on the role of cognitive structures in constructing our experience, that makes a Kuhnian sort of historicism possible. Without Kant, one might be able to trace, as Quine does, the evolution of different ways of thinking about the world. But one would be unable to see, as Kuhn does, the way that historically changing ways of looking at the world actually structure and constrain human experience itself.

In the context of Foucault, too, surrender would require a radical reorientation of Kant's anthropology but need not require rejecting it entirely. Importantly, a strategy like Friedman's – that concedes the historicity of a priori structures of cognition – would not constitute a sufficient concession to Foucault, since Foucault historicizes the whole notion of a priori structures of an individual subject. In the case of Foucault, at least, “surrender” is little more than conversion, a replacement of Kant with Foucault. But arguably, this replacement

remains within the general sphere of Kant. Foucault read Kant's "What is Enlightenment?" as a proto-historicist work (see Foucault 1984: 32-50), and Foucault's intellectual career began with his effort to struggle through problems in Kant's *Anthropology* (see Foucault 2008). Thus Foucault is a sort of radically-historicized Kant, and in that sense, Foucault himself is an excellent model for what a "Kantian-Foucaultian" might look like.

Finally, even a complete Kantian surrender to cultural relativism could still be interesting and important. In particular, Kantian distinctions can help cultural anthropologists more effectively study other cultures. The difference between transcendental and empirical anthropology is itself an important one for cultural anthropologists. It is one thing to describe the way people act and even the way they use normative language, and it is another thing to investigate the way that thoughts and actions appear from-within. The latter task is, of course, much more difficult than the former for an "outsider," and if the from-within perspectives of others are radically incommensurable with our own, it may even be impossible. But clarifying the distinction will force anthropologists to direct attention and articulate their views in more precise ways. Moreover, even within transcendental anthropology, there is an important distinction between the first-order normative claims that one makes from-within and the elucidation of conditions of possibility for those claims. For example, it is one thing to say, from-within, that the boat is moving downstream; it is another to show – as Kant claims to do – that a condition of possibility of the legitimacy of judgments like this is the projection of a category of causality onto the world. The investigation of transcendental conditions of possibility of "alien" ways of thinking and valuing could be an extremely exciting sort of Kantian philosophical anthropology (here using "anthropology" in something like its contemporary sense), a development of *different* transcendental anthropologies (here using "anthropology" in something like its Kantian sense).

V. Conclusion: the problem of normativity

In various ways, historians, historicist philosophers, and anthropologists challenge Kant's anthropology. The last section looked at a series of approaches that Kantians can use to respond to these challenges while preserving, to varying degrees, a distinctively "Kantian" approach to historical and cultural diversity. But regardless of which of these approaches one adopts, there will arise a further important

question, one that mere *descriptions* of human difference and historical change cannot answer: What are the implications of human differences for how we, here and now, *should* think, feel, and act? What is the best response to the diversity? Substantive relativism? Transcendental historicism? Kantian universalism? Something else? As *descriptions* of human difference and historical change, these historicist and anthropological accounts are challenging and illuminating. But in themselves, they don't tell us what to *do* with these descriptions, what effects they should have on how we think and feel and act.

In his *Birth of the Clinic*, Foucault highlights this problem: his method “is concerned – outside of all prescriptive intent – with determining the conditions of possibility of medical experience in modern times” (Foucault 1994: xix, emphasis added). In their pivotal study of Foucault, Hubert Dreyfus and Paul Rabinow add, with respect to Foucault's self-proclaimed goal of “seeking to give new impetus . . . to the undefined work of freedom” (Foucault 1984: 45-6) by resisting the “docile normalcy” of carceral society,

What is wrong with carceral society? Genealogy undermines a stance which opposes it on the grounds of natural law or human dignity . . . Genealogy also undermines opposing carceral society on the basis of subjective preferences and intuitions . . . What are the resources which enable us to sustain a critical stance? (Dreyfus and Rabinow 1982: 206) ²⁵¹

The same point could be made with respect to Kuhn. If the categories and practices of modern science are those of paradigms that emerged recently and may be transcended, what should we do? Should we oppose normal science and try to think of what is truly necessary? Should we simply embrace a scientific process that may be leading us down a misguided path? Similarly, Benedict and Geertz argue that our moral and epistemic values may not be shared by other cultures. If this is true, what should we do? Should we abandon those values? Should we adopt the values of the Dobu? (If so, whose Dobu?) Of the Balinese? (Whose Balinese?) Some new set of values?

These questions are simply a reminder that there is, in fact, a from-within, norm-governed perspective. And they are a reminder that empirical descriptions of the world in which we live – whether these are natural-scientific or historical – are insufficient to answer normative

²⁵¹ Foucault seems to recognize that history provides no guide here: “The historian's essential role . . . [is to provide] a topological and geological survey of the battlefield . . . But as for saying, ‘Here is what you must do!’ , certainly not” (Foucault 1980:62).

questions about how we ought to think, or feel, or act. Foucault, Kuhn, and Benedict show that the fact that one holds a particular normative standard can be explained in terms of historical and cultural conditions. But when I, or you, or Foucault or Benedict, decide whether or not to apply a standard, even one that has been revealed as historically-local, we cannot merely think about that standard in historicist or culturally-relativist terms. While Foucault's histories or Geertz's ethnographies might broaden our range of possible ways of thinking and acting, they cannot in the end tell us how to decide, within that range, what to think or do.

One response to this predicament – Kant's – is to look again, from-within, at the ever-better-understood world in which we live. As we gain new insights into other time periods and other cultures, we can ask ourselves transcendental questions about those insights themselves: what are the conditions of possibility of historical knowledge? How is it possible that we are able to know such things about other cultures? What are the limits of such knowledge-claims? For Kant, the answers to these questions will *include* his general conditions of possibility for knowledge (space, time, causality, etc), and will likely include other more specific conditions of possibility (in the way that biological knowledge requires assuming purposiveness). We can go further, developing our transcendental anthropology of volition in the light of new insights about other cultures. What sorts of obligations might one have towards people with different moral norms? What are the conditions of possibility of mutual respect across cultural difference? What are my duties towards very different others, especially as I come into greater contact with them? And while these questions, for Kant, involve various new subsidiary moral principles, they all require application of the universal moral law.

One might, of course, seek *other* normative responses to the situation of historical and cultural diversity. In chapter eleven, we will look at several such responses, responses that preserve the basic notion that there are norms that can and should govern our thought and action from-within, but that vary regarding the universality and foundations of those norms. One might also – like Nietzsche – take diversity as a basis for liberation into a creativity that rises above present values. The lesson of historicism and cultural diversity, one might think, is that we should stop *looking* for absolute standards “out there,” and start *making* cognitive and volitional standards for ourselves. In its most influential modern form, this emphasis expresses itself in existentialism, to which we now turn.

Chapter 10: Existentialism

In 1966, in a review for the French journal *L'Arc*, Jean-Paul Sartre – who would become the preeminent voice of existentialism for a generation – criticized Foucault's *On the Order of Things* – the book in which he most pointedly discussed Kant's conception of the human being:

What do we find in *The Order of Things*? Certainly not an archaeology of human sciences. Archaeology . . . studies a style that had been designed and implemented by men. This style could thereafter present itself as a natural state, taking the allure of something given. It is nonetheless the result of a practice, the development of which the archaeologist traces. What Foucault offers is . . . a geology: . . . Each layer defines the conditions of possibility of a certain type of thought that triumphs for a certain period. But Foucault does not tell us what is most interesting: how every thought is built from these conditions, nor how people pass from one thought to another. This would require the intervention of praxis, thus history, and this is precisely what he refuses. Certainly its perspective remains historical. It distinguishes the epochs that precede from those that come after. But he replaces the cinema with the magic lantern, the movement with a succession of static states. (Sartre 1966)²⁵²

As we saw in the last chapter, historicists such as Kuhn and Foucault rightly pointed out the role of a “historical a priori,” a set of structures or paradigms that shape the way human beings think about and act within the world. Sartre, in one sense, agrees wholeheartedly with this historicist turn. Like Foucault (and Kuhn), Sartre insists that humans see the world in the context of styles of thinking that structure possible ways of thinking and acting. Like historicists, Sartre sees these a priori structures of cognition as historically contingent rather than universal across all times and peoples. But Sartre, Foucault, and other historicists²⁵³ focus too much on looking at human beings from-without

²⁵² At the time of the review, Foucault had not yet articulated his “genealogical” method, which arguably comes closer to what Sartre sought in that it offers at least some outlines of explanations of the development of different modes of thought. Precisely because these explanations are offered “from-without,” however, and especially given Foucault's radical questioning of the subjectivity that lies at the heart of Sartre's existentialism, Sartre would still see Foucault's account as replacing the cinema (within which subjectivity takes center stage) with the magic lantern (where all changes happen externally). (Though cf. Foucault's *Care of the Self*, where he moves closer to Sartre.)

²⁵³ Kuhn, however, like prior Anglo-American historians of science and unlike Foucault, tends to write a history of science within which individuals loom large. Thus Kuhn provides more room for individual human paradigm choices that are undetermined by “evidence” and even by the social interests and power relations.

and thereby fail to recognize the role of human subjectivity for effecting shifts in the paradigms that structure humans' experiences (including the ways historicists see the world). Similarly, Sartre argues that natural scientists who see human beings in terms of various natural forces (of biology or psychology) fail to recognize the role of subjectivity in defining the meaning and significance of our natural condition. Whereas historicists and naturalists see human beings primarily as the products of historical or natural forces, Sartre insists that history and even biology (insofar as they are significant) are products of humans' responses to their situations. And this opens the way for thinking of paradigms as expressions of human freedom rather than mere forces that constrain us.

Alongside the rise of naturalist and historicist approaches to human beings, the past century has seen the birth and development of "existentialist" approaches to being human that emphasize the perspective from which one sees oneself as a free albeit finite being confronting a world of possibilities. Existentialism has its origins in the 19th century (especially in the work of Kierkegaard²⁵⁴ and Nietzsche²⁵⁵) but came to its own during the 20th century, as the spread of science and technology both radically increased the range of options for human beings and radically narrowed our self-conceptions. The material expansion of choices for many people (especially in the developed world) heightened the sense that who we are is largely up to us. Existentialism rejects naturalist and historicist approaches as the last word on what it means to be human, prioritizes our sense of ourselves from-within, and emphasizes the importance of freedom for human life. The "existential phenomenology" developed in different ways by Heidegger, Sartre, Merleau-Ponty, and others has exerted an important influence not only on contemporary philosophy but on our culture and popular conceptions of what it means to be human.

At the same time, the past twenty or thirty years have seen a turn away from the perceived egocentrism of traditional (especially Sartrean) existentialism. The rise of "deconstructive" approaches to the self, especially in the work of Derrida and Levinas, has combined

²⁵⁴ Or at least, of Kierkegaard's pseudonyms. For the purpose of this chapter, I treat Kierkegaard as the author of his pseudonym's views. For Kierkegaard's own discussion of his relationship with his pseudonyms, see Kierkegaard 1992.

²⁵⁵ One might also include Dostoyevsky here, and both Kierkegaard and Nietzsche are complicated antecedents of existentialism, for different reasons. More generally, with the exception of Sartre, there is no figure who can unambiguously be counted an existentialist, and many who might qualify under one or another description. For some general discussions of existentialism, see the works at the end of this chapter.

existentialist resistance to formulaic reductions of human life with an emphasis on the radical and primordial heteronomy required by one's interactions with "alterity" (the incomprehensible otherness of another). This has brought a shift away from the "ontological" and first-personal perspective of existentialism towards ethical and radically second-personal approaches.

In many respects, existentialism is the most Kantian of the contemporary approaches to the human being that we have discussed so far. Like Kant, existentialists emphasize the importance of freedom and finitude for making sense of being human. And like Kant, existentialists focus on what being human means from-within, rather than analyzing human beings as objects in the world. But existentialists radicalize and modify these Kantian themes. Where Kant defends freedom-as-autonomy subject to a moral law with a determinate form, Sartre insists upon an "absolute freedom" that can appeal to no "book of ethics" (Sartre 1993: 25) and Kierkegaard describes an "absolute duty" that suspends the ethical (Kierkegaard 2006). Where Kant emphasizes finitude primarily in the contexts of sensibility and inclination, existentialists attend in detail to what it means for human beings to have a past, a body, and to be located in a particular situation. Finally, while Kant offers accounts of cognition, feeling, and volition "from-within," these faculties are, at least broadly speaking, taken as self-consciously reflective and highly structured. By contrast, existentialists aim to analyze what they take to be more fundamental, primordial "from-within" perspectives, those of everyday lived experiences of a meaningful world.

To better highlight the significance of these existentialist developments of Kantian themes, this chapter focuses on five key aspects of existentialist thought, drawing primarily from the existentialist philosophies of Sartre and the early Heidegger.

- (1) Human "existence." Existentialists use the notion of "existence" (or "Da-sein" or being "for-itself") to distinguish the way humans exist from the being of things in the world.
- (2) Freedom. As for Kant, focusing on the "from-within" perspective of human being leads to emphasizing human freedom.
- (3) "Being-in-the-world." The from-within perspective of human being is not distinct from practical engagement with concrete situations in the world.

(4) Angst, Bad faith, and Authenticity. Unlike Kantian freedom, which is paradigmatically a freedom to obey moral law, existentialist freedom is groundless. Angst is the experience of this groundless freedom; bad faith and inauthenticity are ways of pretending that we are not really free.

(5) Others. Human being-in-the-world is always also being with- or for-others (other human beings).

From my discussion of existentialist approaches to others (or “the Other”), I turn to the late Heidegger and Levinas (and implicitly also Derrida), who extend existentialist insights but rethink the Other in ways that open space for what has come to be associated with “postmodernism” or “deconstructionism,” but which I describe as “heteronomous” existentialism. I conclude this chapter by discussing ways Kant might appropriate and respond to existentialism.

1. Existence

The name “existentialism” was first used by Gabriel Marcel to describe the circle that grew up around Sartre and Simone de Beauvoir in Paris in the 1940s (See de Beauvoir 1987: 45-6; and Cooper 1990:1).²⁵⁶ The classic formulations of existentialism are Sartre’s claim that “existence precedes essence”²⁵⁷ and Heidegger’s that “The ‘essence’ of Da-sein lies in its existence” (Heidegger 1953:42, see too 117).²⁵⁸ “Existence” contrasts with anything like a human nature (Sartre 1993: 30) or essence that defines the human being. A human being “exists before he can be defined by any concept” (Sartre 1993: 15). What we do with our lives – our “existence” – defines who, and what, we are. Like cultural or historical relativists, existentialists deny that there is any universal answer to the question, “What is the human being?” But existentialists contrast universal human nature not with locally-defined traits but with “subjectivity,” the idea that “Man is nothing else but what

²⁵⁶ Although Sartre’s formulaic claim that existence precedes essence has become existentialism’s most famous articulation, the first to put the term “existence” to use in the way central to existence-ialism was Soren Kierkegaard: “That the knowing spirit is an existing spirit, and that every human being is such a spirit existing for himself, I cannot repeat often enough” (Kierkegaard 1992: 189).

²⁵⁷ “Existentialism is a Humanism,” in *Existentialism and Human Emotions*, p. 13, hereafter abbreviated as E.

²⁵⁸ All quotations from Heidegger’s *Being and Time* are from the translation by Joan Stambaugh. Page numbers are to the seventh German edition.

he makes of himself” (Sartre 1993: 13, 15). The answer to Kant’s question is a matter of how we decide to answer it in our lives. If a human being is anything at all, it is “the being whose being is a question for us” or the being that is always “for-itself.”

Martin Heidegger set the stage for modern existentialism by using Kierkegaard’s concept of “existence” to develop an “existential analysis” (Heidegger 1953:13) of what Heidegger calls “Da-sein.” The German “Dasein” is one of many German words for “existence,” and Germans often refer, for instance, to the Dasein (existence) of a table or chair. But Heidegger uses “Da-sein” in its root meaning of being (Sein) here or there (Da), and he contrasts this sort of Being with the “Being of beings” like tables and chairs.²⁵⁹ One important implication of Heidegger’s reinterpretation of this term is that Da-sein ceases to be a noun (existence) or even an adjective (existent), and becomes, first and foremost, a verb (as in to-be-there, or being-there).²⁶⁰ Like Kierkegaard, Heidegger aims to shift away from thinking of human beings as static objects of study and towards human being, as a sort of activity. The question “What is the human being?” shifts from being about an object – the human being – to being about an adverb: that is, what is a “human” way of being (see BT 45)? What is the sort of “be-ing” that is human?

In order to think about the human be-ing without slipping into forms of thought that have been centered on the analysis of beings (as objects), Heidegger develops a whole new vocabulary of philosophical analysis. His use of Da-sein rather than “human being” is part of this shift, but Heidegger’s new terminology does not end there. Analysis of the “nature” of particular “beings” he calls “ontical,” while the sort of analysis that Heidegger proposes for Da-sein is “ontological,” an analysis of the be-ing of Da-sein rather than an analysis of human beings as objects. And in the context of distinguishing between these sorts of analysis, Heidegger reintroduces the notion of “existence”:

The “essence” of Da-sein lies in its existence. The characteristics to be found in this being are thus not objectively present “attributes” of an objectively present being which has such and such an “outward appearance,” but rather possible ways for it to be, and only this . . . We shall call the characteristics of being of Da-sein existentials. They are to be sharply delimited from the determinations of the being of

²⁵⁹ Though see, for example, Heidegger’s early lectures on Aristotle (in Heidegger 2009), where he moves towards his conception of Dasein by beginning with Aristotelian notions of ordinary objects.

²⁶⁰ This implication does not exhaust the importance of the term Da-sein, for Heidegger. For more on the significance of the “Da-” of “Da-sein,” see section three, below.

those beings unlike Da-sein which we call categories. (Heidegger 1953:42, 44)

While an object might have various properties that inhere in it as a sort of essence, Da-sein has various ways of be-ing as different modes of its own activity. For example, Heidegger contrasts the sort of spatiality that one ascribes to objects, where one might be included in another or beside another, with Da-sein's existential "being-in-space," which is a way humans "be" in the world. Similarly, for Heidegger, a mood like fear is not merely a "state" of a human being, but a way of be-ing (human) (Heidegger 1953:140ff.). More famously, Heidegger discusses death not as a state that brings a particular human being's life to an end, but as an existential "being-toward-death." "Death is a way to be that Da-sein takes over as soon as it is" (Heidegger 1953:245, cf. 236-67). Heidegger's "existential" analyses of Da-sein end up driving him towards complicated German neologisms that require even more complicated English translations, such as his characterization of Da-sein as "being-ahead-of-oneself-already-in (the world) as being-together-with (innerworldly beings encountered)" (Heidegger 1953:192). But for our purposes, the main point of these neologisms is that they reflect Heidegger's efforts to rethink human be-ing using terms that refer to ways of being rather than types of beings.²⁶¹

The emphasis on human being as a way of being has roots in the early 20th century philosophical movement called "phenomenology." Chapter eleven discusses phenomenology as a normative approach, but for this chapter, it is important briefly to discuss it as background to existentialism. Edmund Husserl (1859-1938), who was Heidegger's mentor and heavily influenced Sartre, developed a method for focusing on phenomena of consciousness to gain insight into the essence of conscious experience. Husserl insists on bracketing scientific claims about the world in order to focus on what he calls the "life-world [Lebenswelt]" and to isolate what he calls the "transcendental ego" (Husserl 1975:11). The point, in Kantian terms, is to gain insight into basic structures of naïve (pre-scientific) engagement with the world from-within. While criticizing central details of Husserl's account,²⁶² existentialists in the 20th century maintain this focus on from-within accounts of engagement with our "life-world."

²⁶¹ Sartre takes from Heidegger this emphasis on human being as a way of being but cashes it out in terminology of the "in-itself" and "for-itself." Tables, chairs, waterfalls, and stars are "in-themselves" in that their essence is not an issue for them. But human being is "for-itself" in that "one must be what one is" (Sartre 1956: 101, and passim), that is, what one "is" is a task set for the human being. A table simply is what it is; a human must be what it is.

²⁶² For a concise summary of criticisms, see Cooper 1990: 39-78.

For existentialists, the from-within perspective is privileged in that only for existing human being – what Heidegger calls “Da-sein” and Sartre the “for-itself” – does a world come to “be” at all: “the rise of man . . . causes a world to be discovered” (Sartre 1956: 59, see too Heidegger 1953:12). Existential analysis makes empirical description of human beings secondary. The world is a world only for a human in-the-world. Empirical human sciences are deprivileged; rather than God’s-eye views of human objects, sciences are among ways humans can “be” in the world: “As ways in which human beings behave, sciences have this being’s (the human being’s) kind of being . . . [but s]cientific research is neither the sole nor the most immediate kind of being of this being that is possible” (Heidegger 1953:11). Science is a human practice, a way of existing in the world, so existential analyses of Being-in-the-world explain the possibility of empirical science, rather than vice versa. All of this should sound familiar, and existentialists’ emphasis on existence can be put in rather Kantian terms: Rather than treating humans as objects in the world, existentialists focus attention on what it is like to be human, what living a human life is like from-within.

But existentialists modify this Kantian transcendental perspective in several respects. In contrast to Kant’s transcendental arguments for conditions of possibility of essential ways of human being, the phenomenological method of 20th century existentialists focuses on rich description of what appears or is “disclosed” in humans’ lived experience. For example, Heidegger directs attention to the experience of hammering (Heidegger 1953:69), and Sartre offers detailed existential descriptions of phenomena as diverse as sexual attraction (Sartre 1956: 497f.), shame (Sartre 1956: 350f.), smoking a cigarette (Sartre 1956: 73), and giving in to fatigue during a mountain-climb (Sartre 1956: 584f.). This contrast should not be overdone, of course. Kant was also interested in carefully describing what one finds from-within and existentialists often lay out conditions of possibility of phenomenologically-disclosed structures of human being. Generally, however, Kant is more interested in arguing for certain a priori principles as necessary conditions of possibility of what he takes to be fairly obvious aspects of our “from-within” perspective, while existentialists are more interested in carefully describing that perspective in ways that are often far from obvious.

This difference in method is tied to a difference in what we might call naïveté. Kant takes the basic from-within structures of thought, volition, and even feeling to be transparent to reflection. Proper thought involves justified ascriptions of objective properties and relations amongst objects situated in (Euclidian) space and ((Levinas 1998a:jective) time. Volition involves the pursuit of particular ends by

means of particular actions, and human beings recognize that such pursuits are (morally) justified only insofar as they conform to a standard that is categorical. And so on. But for the existentialists, the basic structures of being human are not at all transparent. The from-within perspectives that become evident upon reflection are not the primary perspectives from within which human beings think, choose, and live our lives. And our naïve conceptions of what it means to think and choose from-within are pervaded by objectivizing and scientific perspectives of which we are often not even aware. As a result, it requires great care and attention to bracket scientific and commonsense prejudices about what it means to “understand” something or to “desire” or “choose” something and instead to genuinely let the structure of human being disclose itself.

2. Freedom

Unsurprisingly, the existentialist emphasis on “existence” as a perspective from within which and acts within and discovers the world leads to an existentialist emphasis on freedom. Sartre puts the point most dramatically, insisting that “existence” is nothing other than “the sudden thrust of the freedom which is mine,” so that “Freedom is identical with my existence” (Sartre 1956: 572). The fundamental statement of existentialism could read, “Human freedom precedes essence in man and makes it possible” (Sartre 1956: 60), or, more paradoxically, “human reality . . . is what it is not and . . . is not what it is” (Sartre 1956: 100 and passim, cf. Heidegger 1953:12). One is defined not by what one “is” at any given time, but by possibilities that one, at present, “is” not but that are nonetheless possibilities of being for one. Similarly, for Heidegger human being is “being-possible” or “potentiality-for-being” (e.g. Heidegger 1953:193): “Da-sein is always what it can be and how it is its possibility” (Heidegger 1953: 143).

The emphasis on freedom is cashed out, in both Heidegger and Sartre, in terms of temporality.²⁶³ Human being is always “being-ahead-of-oneself” (Heidegger 1953:H 192-3), so Da-sein’s Being-in-the-world is always incomplete, or better, to-be-completed. And it is only because “One *must be* what one is” (Sartre 1956: 101, emphasis added) that “I am

²⁶³ Arguably, this is also true for Kierkegaard and Nietzsche. Temporality shows up in Kierkegaard’s distinction between existence as an ongoing process of becoming and the eternal, timeless conceptions of human being that he rejects and in his discussion of the “moment” (see especially Kierkegaard 1983, 1996). Nietzsche’s claim that “your true nature lies, not concealed deep within you, but immeasurably high above you” (Nietzsche 1997: 129) is tied to both a looking back and – more importantly – a projecting forwards into one’s (future) possibilities.

not what I am” (Sartre 1956: 108). That is, freedom is found in the fact that human being is defined in terms of its future. Because Da-sein is always a potentiality for being, there is an essential “not-yet” associated with Da-sein. Sartre gives the example of writing a book:

I cannot write words without transcending them toward the sentence [I]n every act of this kind there remains the possibility of putting the act in question insofar as it refers to more distant . . . ends . . . [T]he sentence . . . is the meaning of the letters . . . but the whole work . . . is the meaning of the sentence. And this work is a possibility in connection with which I can feel angst; it is truly *my* possibility. (Sartre 1956: 74)

This future-orientation brings with it an essential “being-ahead-of-itself” (Heidegger 1953:236). One’s present and even past are always defined in terms of one’s future; by projecting into one’s future, one makes sense of the past²⁶⁴ and defines the present. But then one’s present and past are always, in some sense, defined by freedom. Only what one will do can define the meaning of what one is doing and has done. As Heidegger puts it, deliberately emphasizing the existentialist inversion of typical conceptions of time, “having-been arises from the future” (Heidegger 1953:326).

Importantly, existentialist freedom is not limited to freedom of action: “the *world* is human” (Sartre 1956: 297).²⁶⁵ More precisely,

The world by means of its very articulation refers to us exactly the image of what we are . . . the world necessarily appears to us as we are. In fact it is by surpassing the world toward ourselves that we make it appear such as it is. We choose the world, not in its contexture as in itself but in its meaning, by choosing ourselves. (Sartre 1956: 596)

Being is what it is; it can not possess in itself the determination “this one” . . . it is the presence of the for-itself which causes the existence of a “this” rather than a “that.” . . . Negativity as original

²⁶⁴ Sartre puts this point in an even stronger way. It is by virtue of projecting oneself into one’s future that there is a past at all: “For human reality alone the existence of a past is manifest because it has been established that human reality has to be what it is. It is through the for-itself that the past arrives in the world because its “I am” is in the form of an I am me.” (Sartre 1956: 168, cf. Sartre 1938)

²⁶⁵ Kierkegaard emphasizes this epistemic existentialism (see, e.g., Kierkegaard 1992: 189).

transcendence is not determined in terms of a this; it causes a this to exist. (Sartre 1956: 249, see too 264)

The brute world as it might be considered in itself is wholly undifferentiated, but the world that we actually know and experience is always already structured in terms of our own projects and priorities. Thus our knowledge, and even the world itself that we know, follow from our free orientation towards that world. As Heidegger puts it,

“There is” truth only insofar as Da-sein is and as long as it is . . . Newton’s laws, the law of contradiction, and any truth whatever, are true only as long as Da-sein is . . . If no Da-sein exists, no world is “there” either. (Heidegger 1953:226, 365)²⁶⁶

Put another way, “It is freedom which is the foundation of all essences since man reveals intra-mundane essences by surpassing the world towards its own possibilities” (Sartre 1956: 567).

This emphasis on freedom sets the existentialists with Kant and against naturalists and historicists. For existentialists, freedom does is not merely something that emerges from a naturalistically or historically determined world, but that by virtue of which there is history or a world at all.²⁶⁷ But existentialists’ emphasis on freedom diverges from and challenges Kant’s transcendental anthropology in two ways. First, existentialist freedom is always freedom in-the-world. As Sartre puts it, “we reject Kant’s ‘choice of intelligible character.’ The structure of choice necessarily implies that it be a choice in the world” (Sartre 1956:, 617). Similarly Heidegger emphasizes that human possibility is always “thrown possibility” (Heidegger 1953:144):

As an existential, possibility does not refer to a free-floating potentiality of being in the sense of the “liberty of indifference” (*libertas indifferentiae*). As essentially attuned, Da-sein has always already got itself into definite possibilities. (Heidegger 1953:144).

For existentialists, Being-in-the-world is inseparable from freedom and at least as central as freedom to understanding “existence.” Second, existentialists reject Kant’s conception of freedom as autonomy, Kant’s notion that to choose freely is to subordinate one’s will to a universally applicable categorical imperative, or that to think spontaneously is to

²⁶⁶ Heidegger rightly notes that in this sense, “as opposed to realism, idealism . . . has a fundamental priority” (Heidegger 1953: 207).

²⁶⁷ In terms of the conceptions of freedom laid out in chapter eight, existentialists share with Kant a commitment to freedom-first perspectivism (see p. xxx).

think in accordance with universally valid a priori categories. Instead, existentialists insist upon what Sartre calls “absolute freedom” (Sartre 1953: 653), a freedom that claims no justification; and existentialists emphasize, against the Kantian ideal of autonomy, an ideal of authenticity according to which one pursues one’s “ownmost” possibilities.

3. Being-in-the-world

The existentialist emphasis on Being-in-the-world is one of its most important challenges to Kant. For Heidegger, the “being which we ourselves in each case are” is not merely “Sein,” or “Being,” but Da-sein, or Being-there. Human being is always situated in a particular context. Whereas Kant generally treats the from-within perspective as a sort of “view from nowhere” (albeit a distinctively human one),²⁶⁸ existentialists emphasize that human existence is always existence somewhere, in a specific situation and with concrete projects and possibilities. Take the examples Kant uses to establish the a priori principle of causation: the house and the boat (see chapter one). Kant imagines how disinterested and only barely embodied observers would collect a set of subjective perceptions of boat-states or house-states into coherent objective states of co-existence or succession. But he does not think about where in this situation the knower finds herself, nor why she is looking at the house (or boat), nor how she came to find herself seeking to construe these objects as objects. Kant’s knower could be anyone anywhere with no particular stakes in the situations she surveys. Existentialists, by contrast, emphasize that, even from-within, human being is not a disembodied contemplation of the world, but a sort of being that always finds itself already “there,” or, as Heidegger puts it more precisely, always “in-the-world.” The person surveying the house is not only located some distance from it, but the survey of this house is conducted for some reason – the onlooker may be a prospective buyer, or may be coming home, or approaching a strange house to ask directions – and everything about the house (including, for instance, its perceived “distance” from one and even the detail with which one perceives it) is affected by one’s purposes in surveying it.

Heidegger reflects this difference between Kantian “experience” and existentialist Being-in-the-world in his account of the paradigmatically human interaction with the world we experience. Heidegger criticizes

²⁶⁸ See Nagel 1986, but cf. Hanna 2006.

Kant (and the whole Western metaphysical tradition) for focusing on objects that are “conspicuous” or “merely objectively present” (Heidegger 1953:71, 73), that is, objects that are the direct focus of reflective understanding. Heidegger points out that the beings we encounter are initially and for the most part “handy” objects, that is, beings that we put to use:

The closest kind of association [with things in the world] is not mere perceptual cognition, but, rather, a handling, using, and taking care of things which has its own kind of knowledge . . . To expose what is merely objectively present, cognition must first penetrate beyond things at hand being taken care of. Handiness is the ontological categorial definition of beings as they are “in themselves.” (Heidegger 1953:67, 71)

Kant’s house and boat are not first objects to be surveyed, but places to live or vehicles for travel on the river. The objectivizing reflection Kant takes as fundamental is actually a highly derivative form of knowing that abstracts from lived knowledge of the world of handy objects that always already have meaning in the context of Da-sein’s being-in-the-world. Because human being is this sort of “Being-in-the-world” rather than being-from-nowhere, human thoughts and actions cannot be separated from their contexts. Sartre describes this as our “facticity” (Sartre 1956 passim, cf. 617, 669) and Heidegger as what he calls “thrownness” or our “thrown possibility” into “definite possibilities” (Heidegger 1953:144).²⁶⁹

Being-in-the-world has radical implications for existentialists’ conceptions of human freedom. Unlike naturalists, existentialists do not take freedom to be a mere epiphenomenon of what is a fundamental causal determination. Existentialist freedom – like Kant’s – determines rather than being determined by the world. But this freedom finds itself always already thrown into its situation. Sartre describes this as “the paradox of freedom: there is freedom only in a situation, and there is a situation only through freedom” (Sartre 1956: 629, see too Sartre 1993: 23, Heidegger 1953:299-300), such that “I am absolutely free and absolutely responsible for my situation. But I am never free except in situation” (Sartre 1956: 653). Or, as he puts it elsewhere, “Human reality

²⁶⁹ Thus Sartre suggests that asking, “what would Descartes have been if he had known of contemporary physics?” is “absurd” precisely by failing to recognize that Descartes is always Being-in-the-world. As Sartre puts it,

This [question] . . . suppose[s] that Descartes possesses an a priori nature more or less limited and altered by the state of science in his time . . . [But in fact] Descartes is an absolute upsurge at an absolute date and is perfectly unthinkable at another date, for he has made his date by making himself.” (Sartre 1956: 669)

is indeed the being which is always beyond its being-there. And the situation is the organized totality of the being-there, interpreted and lived in and through being-beyond" (Sartre 1956: 702). What is, "is-there" only by being interpreted in terms of projects and possibilities towards which I freely orient myself. But I can interpret as meaningful only a world that I find, at least in some sense, always already there as that beyond which I find my possibilities. Material wealth, for example, only becomes what it is – a means for luxurious living, a source of power over others, a temptation to impious self-reliance, a burden of responsibility – in terms of projects freely projected. Even whether or not one is materially wealthy depends upon one's projects and possibilities: neither the sheltered prince (who knows no other life) nor the ever-grasping corporate executive (who sees his six-figure income only in terms of the seven-figure income that it is not) sees their situation as one of wealth. Yet neither of these figures is capable of defining their situation in terms of the struggle for the means for physical survival. The freedom of each is constrained by their situation, but the meaning of this situation is itself defined by them.

In addition to refining the existentialist conception of freedom, understanding human being as "Being-in-the-world" also leads existentialists to reject a host of dualisms that permeate Kant's philosophy. The dualism on which existentialists themselves typically focus is that between subject and object, which Sartre describes as "useless and disastrous" (Sartre 1956: 318, see too Heidegger 1953:176). In fact, however, the existentialist rejection of this dualism is more subtle than existentialists often admit. Both Heidegger and Sartre have distinctions, if not dualisms, that resemble Kant's distinction between transcendental subject and worldly objects. For all his insistence that "subject and object are not the same as Da-sein and world" (Heidegger 1953:H 60), Heidegger's distinction between Da-sein and mere beings and Sartre's between the "in-itself" and the "for-itself" are attempts to lay out the difference between free, from-within, human being and the being of the objects of human thought and action.

The vitriol against Kant is really directed towards two problems with how existentialists see Kant making sense of the distinction between transcendental and empirical perspectives. One problem is simply that Kant fails to sufficiently recognize that human being is always being-in-the-world. By positing a subject-object dichotomy, or a distinction between the self-in-itself and the realm of appearance, Kant fails to recognize that "it belongs to the nature of Da-sein to exist in such a way

that it is always already with other beings” (Heidegger 1982: 157, quoted in Cooper 1990:81). As a criticism of Kant’s subject-object dichotomy, this objection ignores Kant’s own emphasis on humans’ finitude. While Kant rejects aspects of existentialist being-in-the-world, the basic notion that human cognition, feeling, and volition take place in the context of situations in which one always already finds oneself is intrinsic to the balance between freedom and finitude that lies at the core of Kant’s transcendental anthropology.

A second problem with the subject-object dichotomy is that laying out the dichotomy in these terms makes free human being too much like the being of objects. The danger here is that this distinction ends up construing subject and object as, in essence, two different sorts of objects that relate to one another. Here, existentialists highlight an important difference from – and potential problem with – Kant’s account. Strictly speaking, Kantian “objects” are empirical, subject to the transcendental conditions of experience (space, time, and categories like causation), and thus the human being as subject of transcendental anthropology cannot be an “object.” Nonetheless, for Kant it makes sense to talk about *the* human being (as a sort of thing), rather than human being (as a way of being). And Kant’s accounts of “intelligible character” and the homo noumenon suggest what has been called a “two-world” account in which human beings exist as free things-in-themselves in a “noumenal” world while ordinary objects, including the appearances of human beings in the empirical world, exist in a “phenomenal” world.

In the context of this interpretation of Kant’s transcendental idealism, the existentialist critique of Kant reconnects with criticisms raised against Kant in the 18th and early 19th century. Echoing Jacobi’s “Affection Problem,” Sartre points out that “a priori th[e] spontaneity [of the noumenal free self] would be capable of no action on a determinism [of phenomenal actions] already constituted” (Sartre 1956: 570). More generally, insofar as the notion of “the” human being as a “thing-in-itself” that grounds the world involves the attribution of merely empirical categories to a non-empirical thing, existentialists are correct that Kant cannot afford to keep this notion. But within Kant’s transcendental anthropology of cognition, categories of the understanding provide for the possibility of “thinking” of things that cannot be objects of any possible experience, and his practical philosophy provides reason to believe in the existence of at least one such thing – the free human being. In the context of the existentialist critique, Kant needs at least to show how thinking of human beings as things-in-themselves is consistent with the non-objective nature of human being. He further needs to alleviate concerns that this sort of dualism precludes the integration of freedom

and world necessary to capture the most important existentialist insights about being-in-the-world. And he would need to provide strong reasons to think that the distinction between transcendental and empirical perspectives, so important for both Kant and existentialism, actually depends upon metaphysical commitments to separate realms within which human beings exist. Some contemporary Kantians have sought to meet these existentialist challenges. Others – we will discuss one prominent example in the next chapter – articulate accounts of Kant’s philosophy that construe Kant’s transcendental idealism in a broadly existentialist way (and thus as not involving any commitment to the metaphysical reality of “things-in-themselves”).

4. Being-in-the-world and the status of science.

One might think that understanding human being as “being-in-the-world” would privilege natural sciences, but existentialism implies precisely the opposite. For existentialists, being-in-the-world is not an attribute of an object; humans are not like trees or spoons that are “in-the-world” in the sense of being spatially located in a greater context. For us, “being-in-the-world” is a way that we are from-within. Moreover, being-in-the-world, precisely because it collapses Kantian (and modern scientific) distinctions between knowing, feeling, and acting, undermines sciences’ supposed “objectivity.” Heidegger and Sartre both discuss naturalism in detail, and both seek to show how a purportedly “objective” perspective can arrive, as a “way in which human beings behave” (Heidegger 1953:H 11), from our more basic engagement with the world: “Even the disinterested attitude of a scientist . . . is the *assumption* of a disinterested position with regard to the object and hence one conduct among others” (Sartre 1956:, 613).

Existentialists end up sufficiently more anti-naturalist than Kant that they challenge Kant’s limited scientific realism. For Kant, although the natural sciences do not reveal everything about everything, they provide our best account of the world we experience. By collapsing the distinction between practical projects and empirical cognition of the world, existentialists undermine Kant’s attempt to carve out this privileged space for science. Sartre puts the point particularly dramatically:

I can establish that the warm water appears cold to me when I put my hand in it after having first plunged my hand in hot water. But this establishment which we pompously call “the law of relativity of

sensations” has nothing to do with sensations. Actually we are dealing with a quality of the object which is revealed to me: the warm water is cold when I submerge my heated hand in it. A comparison of this objective quality of the water to equally objective information which the thermometer gives me simply reveals to me a contradiction. This contradiction motivates on my part a free choice of true objectivity. I shall give the name subjectivity to the objectivity which I have not chosen. (Sartre 1956: 412)

Where Kant privileges the scientific standpoint as the correct and normatively required standpoint for understanding the world (even if not the exhaustive standpoint on human action), Sartre sees merely various sets of incompatible objectivity-claims, the priority of which is a matter of “free choice.” Moreover, existentialists’ emphasis on the “life-world” suggests that the immediate awareness of the water as cold is the more important, while science is secondary, derived, and, in its theory-ladenness, subjective. As Heidegger puts it with respect to space and time,

An ‘objectively’ long path can be shorter than an ‘objectively’ much shorter path which is perhaps an ‘onerous one’ . . . When there is a prior orientation towards [scientific] ‘nature’ and the ‘objectively’ measured distances of things, one is inclined to consider such interpretations . . . ‘subjective.’ However, that is a ‘subjectivity’ which perhaps discovers what is most real about the ‘reality’ of the world. (Heidegger 1953:106).

One might see the emphasis on life-world as precisely what Kant needs to salvage his transcendental anthropology in the light of the scientific historicism discussed in the last chapter (Kuhn). As I suggested there, one might read Kant’s transcendental anthropology of cognition as an account of ordinary human experience and thus insulate it from criticisms based on the development of the sciences. Unfortunately, existentialists show that, once the distinction between thought and action is rejected, once human “understanding” is disclosed as a way that “Da-sein projects its being upon possibilities” (Heidegger 1953:H 148), Kant’s most basic analyses of human experience fail to reflect what ordinary life is really like “from-within.” Just as Kuhn seems to undermine Kant’s account of the conditions of possibility of scientific knowing, Heidegger and Sartre aim to undermine Kant’s account of the conditions of possibility of ordinary, lived, human experience.

At a very basic level, this is accomplished through drawing attention to the priority of engaging with things as handy rather than as

“objects” of theoretical knowing. But the existentialist critique cuts deeper as it gets more specific. One of the most important examples of this undermining of Kant’s transcendental philosophy is the existentialist re-interpretation of the nature of time. (Heidegger’s magnum opus is not called *Being and Time* by accident.) Heidegger restores a focus on lived time and points out that, within lived experience, time does not appear as what Kant called “objective succession.” Within “time” as objective succession, even the most basic temporal categories of past, present, and future are merely derived relations between the time in which events take place and the temporal location of oneself. The categories of “past, present, and future” are all-too-commonly used in a sense derived from the broadly Kantian, “vulgar . . . derivative . . . [and] inauthentic” conception of objective time, where these are merely objectively-locatable time-slices that include different moments. Instead, Heidegger develops a notion of “temporality” as “the unified phenomenon of the future that makes present in the process of having-been” (Heidegger 1953:H326). Heidegger’s complex terminology draws attention to an important point about our lived experience of temporality. From-within, the “present” does not appear as a particular moment between other moments, some past and some future. Instead, human existence is a projecting towards possibilities, a projecting that is always in the process of having-been in-the-world that one always-already is. The future is not first and foremost an objective time that will come to be present later, but a set of possibilities towards which one aims. The past is not first and foremost a set of events that occurred in previous moments, but rather that in one’s situation that is given, that one can (re)interpret but not “change.” And the present is not a moment that happens to occur between past and future but rather it is one’s existence itself, that way of being from the past into the future. Within our ordinary lived experience of the world, the heat-death of the universe is not an event in our future and the Big Bang not an event in our past; these moments do not show up in “our” time at all. And the derision that “will” be directed towards me as a result of the obnoxious email I overhastily sent is as much a part of my “past” as the event of clicking “send.” (I act towards my future possibilities in the light of an embarrassment or defensiveness that reflects the fact that this derision has occurred, even though, in terms of “objective” time, the recipient of my email “will” not open it until tomorrow morning.)

What Kant isolates as “time” is thus not the lived temporality of human being but a highly derived sense of “objective” time, a sort of “time” that is important for objective, scientific knowledge, but not the primordial temporality of Da-sein. Similar existential reinterpretations could be offered for basic Kantian categories like space, causality, or

substantiality.²⁷⁰ These existential critiques highlight the distance of Kant's transcendental anthropology of cognition from the ordinary ways of cognizing objects in which humans initially and for the most part engage. The "experience" of Kant's first *Critique* already involves an artificial, scientific objectivizing of the empirical world.

5. Absolute Freedom: Angst, Bad Faith, and Authenticity

The consideration with which we ended the last section draws attention to an important way that the limitation on freedom implied by Being-in-the-world also supports a radicalization of freedom. Because human knowing is always situated in the context of our practical projects, the scientific perspective cannot be privileged, and thus it is always, in some sense, up to us whether we employ, say, Kant's a priori categories of the understanding or not. Similarly, because humans always choose in particular contexts, existentialists argue, no abstract moral formulae can dictate how one must choose. Thus while freedom is always situated, it is also more radically free than Kant envisioned because bound by no a priori laws of understanding or volition.

Sartre describes what has become the most famous illustration of this radical freedom, the case of one of his students:

his father was on bad terms with his mother, and, moreover, was inclined to be a collaborationist [with the Nazis]; his older brother had been killed in the German offensive of 1940, and the young man, with somewhat immature but generous feelings, wanted to avenge him. His mother lived alone with him, very much upset by the half-treason of her husband and the death of her older son; the boy was her only consolation.

The boy was faced with the choice of leaving for England and joining the Free French Forces – that is, leaving his mother behind – or remaining with his mother and helping her to carry on. He was fully aware that the woman lived only for him and that his going off – and perhaps his death – would plunge her into despair. He was also aware that every act that he did for his mother's sake was a sure thing, in the sense that it was helping her to carry on, whereas every

²⁷⁰ While Heidegger focuses on human temporality, he offers a similar existential reinterpretation of space as "the whereto of the possible belonging somewhere of useful things at hand in the surrounding world" (Heidegger 1953: 368). Euclidian space, rather than being primitive to our experience, is a sort of abstraction from this lived spatiality.

effort he made toward going off and fighting was an uncertain move which might run aground and prove completely useless; for example, on his way to England he might, while passing through Spain, be detained indefinitely in a Spanish camp; he might reach England or Algiers and be stuck in an office at a desk job. As a result, he was faced with two very different kinds of action: one, concrete, immediate, but concerning only one individual; the other concerned an incomparably vaster group, a national collectivity, but for that very reason was dubious, and might be interrupted en route. And, at the same time, he was wavering between two kinds of ethics. On the one hand, an ethics of sympathy, of personal devotion; on the other, a broader ethics, but one whose efficacy was more dubious. He had to choose between the two.

Who could help him choose? . . . No book of ethics can tell him. The Kantian ethics says, "Never treat any person as a means, but as an end." Very well, if I stay with my mother, I'll treat her as an end and not as a means; but by virtue of this very fact, I'm running the risk of treating the people around me, who are fighting, as means; and, conversely, if I go to join those who are fighting, I'll be treating them as an end, and, by doing that, I run the risk of treating my mother as a means. If values are vague, and if they are always too broad for the concrete and specific case that we are considering, the only thing left for us is to trust our instincts. That's what this young man tried to do; and when I saw him, he said, "In the end, feeling is what counts. I ought to choose whichever pushes me in one direction. If I feel that I love my mother enough to sacrifice everything else for her – my desire for vengeance, for action, for adventure – then I'll stay with her. If, on the contrary, I feel that my love for my mother isn't enough, I'll leave. (Sartre 1993: 24-26)

For Sartre, this case illustrates how no particular ethical system antecedently determines what I ought to do: "we apprehend our choice—i.e., ourselves—as unjustifiable. This means that we apprehend our choice as not deriving from any prior reality" (Sartre 1956: 598).

Similarly, Heidegger discusses Kant's "fact of reason" as a "fact . . . always and only given by us to ourselves," but Heidegger argues against any attempt to understand this fact in terms of "any formula or . . . value held up before us" (Heidegger 2002: 201). Instead, the fact of pure reason is given only in our "resolve to pure willing or against this," where pure willing is "to be in the mode of self-responsibility, to answer only to the essence of one's self" (Heidegger 2002: 201). Unsurprisingly, "conscience" for Heidegger is not a vague understanding of a possibly articulable

moral law (as in Kant), but rather an invitation to a “resolution” that has no articulable form: “to what does Da-sein resolve itself in resoluteness? On what is it to resolve? Only the resolution itself can answer this The indefiniteness that characterizes every factually projected potentiality-of-being of Da-sein belongs necessarily to resoluteness” (Heidegger 1953:298). Kierkegaard and Nietzsche arguably put this freedom from determinate moral laws in its most radical forms, with Kierkegaard defending a “teleological suspension of the ethical” in the name of an inarticulable and unjustifiable “absolute duty” (Kierkegaard 2006) while Nietzsche proposes moving “beyond good and evil” (Nietzsche 1966) towards “your true nature [that] lies . . . immeasurably high above you” (Nietzsche 1997:129).²⁷¹

Given their rejection of Kant’s distinctions between cognition, feeling, and volition, existentialists also radicalize Kant’s conception of freedom in another way. Although Kant emphasizes the spontaneity of the understanding and the freedom of aesthetic pleasure, volitional freedom is distinguished from and prioritized over these other sorts of freedom. For existentialists, however, the freedoms of thought, feeling, and choice all run together into a single, free, projection towards one’s own possibilities. While Kant insists that we cannot be held directly responsible for emotions because they are not under our direct control (see 5:83), Sartre seeks an “existential psychoanalysis” that ascribes all thoughts, feelings, and deliberate choices to an “original choice” or “fundamental project” (Sartre 1956: 728, 729), “the free project of the unique person” (Sartre 1956: 782).²⁷²

[Existential] psychoanalysis must . . . understand someday why Pierre likes oranges and has a horror of water, why he gladly eats tomatoes and has a horror of beans. (Sartre 1956: 770)

[T]astes do not remain irreducible givens; if one knows how to question them, they reveal to us the fundamental projects of the

²⁷¹ As noted in chapter seven, as a metaphysical matter, Nietzsche sees Kant’s transcendental freedom of the will as a “rape and perversion of logic” (Nietzsche BGE § 21, 1966: 28). But in his practical emphasis on a freedom that goes beyond good and evil, he is in good company with Kierkegaard and Heidegger.

²⁷² This emphasis on choice can easily be misread (as, for example, Merleau-Ponty does in Merleau-Ponty 1945) if one forgets that Sartre (like Kierkegaard, Nietzsche, Heidegger, and others) rejects Kant’s distinctions between cognition, feeling, and volition. Thus the free “choice” to which Sartre refers here is elsewhere called a person’s “passion” (Sartre 1956: 797), and Sartre insists that “the will, far from being the unique or at least the privileged manifestation of freedom, actually . . . must presuppose the foundation of an original freedom” (Sartre 1956: 571).

person. Down to even our alimentary preferences they all have a meaning. (Sartre 1956: 783)

Kant always distinguishes between what is merely given and how one freely responds to what is given, but for Sartre, everything about a person is ascribable to freedom, including not merely choices for or against the moral law, but even apparently instinctual desires, immediate sensory attention to one's world, and apparently uncontrolled "passions" or "moods."

The rejection of Kant's distinction between cognition, feeling, and volition fostered attentiveness to the disclosive function of moods and emotions, among which one mood preeminently discloses humans' absolute freedom: anxiety, or *Angst*.²⁷³ *Angst* has as its object "being-in-the-world as such," and in particular the realization that "the world has the character of complete insignificance" (Heidegger 1953:186). "What oppresses," in *angst*, "is not this or that, nor is it everything objectively present together as a sum, but the possibility of things at hand in general, that is, the world itself" (Heidegger 1953:187). But the world experienced in *angst* is not merely a world set apart from oneself, but the world in relation to "the ownmost individualized being [the "being-possible"] of *Da-sein*" (Heidegger 1953:265, cf. 188). In recognizing that the world is meaningful only through one's own unjustified and unjustifiable being-in-the-world, one experiences the disorientation of freedom.

In *Angst*, . . . we apprehend our choice—i.e., ourselves—as unjustifiable. This means that we apprehend our choice as not deriving from any prior reality. (Sartre 1956: 598).

There is ethical *Angst* when I consider myself in my original relation to values . . . my freedom is the unique foundation of values and . . . nothing, absolutely nothing, justifies me in adopting this or that particular value, this or that scale of values. (Sartre 1956: 76, cf. 65, 70)

Angst is contrasted with mere "fear." Fear is directed towards some threat to one's being, while *Angst* focuses on freedom itself. Sartre explains this point with the example of walking along a narrow precipice. One can be afraid of the precipice: it "presents itself to me as to be

²⁷³ The word *Angst* is from Heidegger's German, and can be translated into English as anxiety or anguish. Since *Angst* has come to be an accepted English word associated specifically with existentialism, I use the term throughout this section, and in my quotations of Sartre, I change Barnes' translations of Sartre's "*l'angoisse*" from "anguish" to "angst."

avoided; it represents a danger of death” (Sartre 1956: 66). But one experiences Angst in relation to oneself and the ever present “possibility . . . to throw myself over the precipice” (Sartre 1956: 67). The recognition that such a course of action is wholly up to me, that there is nothing – no prior commitments, no moral codes, no inclinations or desires – that can prevent me from making the fatal choice—this recognition prompts existential Angst. Similarly, the mid-career professional may suddenly have her angst-filled “midlife crisis” when she realizes that none of her past choices and commitments requires or justifies her present life.

Importantly, Angst does not reveal simply that one is radically free. It also highlights that one’s choices are always one’s own. Insofar as one sees the possibility of falling off a cliff as something that can happen to one, it causes merely fear. But when one sees it as something one can *do*, it is a source of Angst. For Heidegger, the revelation of possibilities as one’s ownmost is preeminently accessible through what he calls “being-toward-death,” and thus “Being-toward-death,” for Heidegger “is essentially Angst” (Heidegger 1953:266). Death, properly understood, “is always just one’s own” (Heidegger 1953:265). While people can think of death merely “objectively,” as an event that will occur *to* one, just as to others, such accounts of death always fails to really understand it. When “one” finishes a project, or insults a person, or even has a child, these are events that could occur in the life of another as much as in one’s own life. In principle, another can even undergo these activities for me: another can finish my project, or insult my enemy, or even have a child on my behalf (either giving birth to one for me or taking responsibility for the one I’ve begotten). But

no one can take the other’s dying away from him. Someone can go “to his death for another.” However, that always means to sacrifice oneself for another “in a definite manner.” Such dying for . . . can never, however, mean that the other has thus had his death in the least taken away. Every *Da-sein* must itself actually take dying upon itself. Insofar as it “is,” death is always essentially my own . . . Death does not just “belong” in an undifferentiated way to one’s own *Da-sein*, but it lays claim on it as something individual (Heidegger 1953:240, 263).

To me, my death cannot be something that another can undertake. If a project in which I am interested gets finished, or if an insult is delivered, or even if a child is successfully raised, I can, at least in principle, take the same stance towards those events whether or not I have performed them. But if “a person” dies, it makes all the difference in the world whether or not that “person” is me. For me, whether death “happens” to

me or to another makes the difference between there being a world, between my being-in-the-world, or not. In that sense, being-toward-death is being-in-the-world with the recognition that one's being-in-the-world is one's own, that one "is" in the world as one's ownmost possibility towards the world. Angst as the revelation of being-toward-death is the revelation of this ownmost possibility.

Famously, Sartre rejects Heidegger's emphasis on death, arguing that in exactly the same sense that no one can die for me, no one can love for me, or do or feel anything for me (Sartre 1956: 684). And my death, for Sartre, is precisely something that cannot be my "ownmost," in that death always "overtakes me" (Sartre 1956: 683).²⁷⁴ Despite this apparently vehement disagreement, however, Sartre and Heidegger fundamentally agree about the underlying point. Heidegger (like Sartre) does not see the state of having-died as something towards which it makes sense to aim in the way one might make having-married an aim. Death is existentially significant as a "possibility" that "gives Da-sein nothing . . . which it itself could be as something real" (Heidegger 1953:262).²⁷⁵ Moreover, like Sartre, Heidegger sees "brooding over death" and even "expecting it" as inauthentic ways to avoid coming to terms with being-toward-death. Moreover, Heidegger agrees with Sartre that being-in-love or any other activity can be just as authentic as being-toward-death insofar as one takes up these ways of being as one's ownmost possibilities. Death reveals one's non-substitutability, but authentic being-towards-death involves treating all one's possibilities as just as non-substitutable as death.

Of course, existential Angst is not the prevailing mood governing most of our lives: "real' Angst is rare" (Heidegger 1953:190). The professional with a mid-life crisis leaps into work even more firmly, buys a zippy sports car, and/or has a fling with a young co-worker. The hiker along the precipice may briefly entertain leaping from it, but quickly recalls obligations to family or expectations of pleasures with friends and focuses on the task at hand. Most of us, most of the time, evade angst by

²⁷⁴ Mocking the idea of being-towards-death, Sartre writes:

It has often been said that we are in the situation of a condemned man among other condemned men who is ignorant of the day of his execution but who sees each day that his fellow prisoners are being executed. This is not wholly exact. We ought rather to compare ourselves to a man condemned to death who is bravely preparing himself for the ultimate penalty, who is doing everything possible to make a good showing on the scaffold, and who meanwhile is carried off by a flu epidemic. (Sartre 1956: 683)

²⁷⁵ Because Heidegger agrees with Sartre that death is not something towards which one can aim as a concrete project, he also rejects suicide as an appropriate (general) existential response to absolute human freedom. See Heidegger 1953: 261, Sartre 1956:690.

focusing on concrete tasks that we assume as given, doing “what must be done,” “what is expected,” or “what I feel like.” Faced with absolute freedom, the typical response, one that for Sartre is all-too-common and for Heidegger is inevitable, is to flee from one’s “ownmost possibilities” into something more comfortable. Heidegger describes this flight as “inauthenticity” (Heidegger 1953) Sartre as “bad faith” (Sartre 1956: 711).

Bad faith is the denial of the fundamental existential insight that “I am not what I am” (Sartre 1956: 108). “What I am” refers to my facticity or being-in-the-world. Freedom starts in a situation defined by choices already made, a past already lived, roles in which we find ourselves. “I am not” refers to the always present absolute freedom. Whatever “I” am in my facticity – a father, a writer, a lover of chocolate, a philosopher – I am not that facticity. I can choose to accept or reject any particular role, any particular desire, any particular past. Or better, by taking up my situation, my past, my roles in terms of different future possibilities, I define the facticity that “I” am, and thus I am not reducible to that facticity. Bad faith denies one or another of the aspects of this “I am not what I am.” We might take, as an example, someone who avoids socializing and is frequently shy or awkward in social situations.²⁷⁶ Such a person might say of herself, perhaps in response to criticism, “I’m not shy.” Sartre explains,

She would actually be right if she understood the phrase, “I am not shy” in the sense of “I am not what I am.” That is, if she declared to herself, “to the extent that a pattern of conduct is defined as the conduct of someone who is shy, and to the extent that I have adopted this conduct, I am shy. But to the extent that human reality can not be finally defined by patterns of conduct, I am not shy.” [But insofar as] . . . she lays claim to “not being shy” in the sense in which this table is not an inkwell, she is in bad faith. (Sartre 1956: 108)

When the girl says “I am not shy” in order to deny her facticity, she speaks in bad faith; but were she to say “I am not shy” in order to deny that she *is* her facticity, she would speak truly. Alternatively, our shy character might affirm, “I am shy.” Here, too, this is true only in the sense that “I am not what I am.” This self-understanding is authentic only insofar as she understands shyness as something for which she is responsible, something up to her. Insofar as she claims to “be shy” in the sense that the table “is a table,” she is, again, in bad faith. In each case,

²⁷⁶ Sartre gives, as an example, a “homosexual” (who he also refers to as a “paederast”, Sartre 1956: 107-8) but we could envision any character who is ashamed of past behavior or of his or her desires, preferences, etc. In the following quotation, I’ve modified the gender of the character and the relevant characteristic (from pederasty to shyness).

she ignores either her facticity, supposing that freedom takes place without situation, or her freedom from her facticity, supposing that previous patterns of behavior wholly determine who she is. Both poles are forms of denying one's absolute and radical freedom, a freedom that is the source of one's situation even as it transcends that situation.

Rather than living in bad faith, one can and should live in good faith, "authentically" or "sincerely." Authentic living requires recognizing that one's choices are one's own. It requires taking up one's facticity in the light of the free projection of possibilities. Sartre explains, "One may choose anything if it is on the grounds of free involvement" (Sartre 1993: 47-8). But choosing in a way that is really free, and really involved, is not as straightforward as it sounds. It requires, as Heidegger puts it, heeding the "call of conscience" that "*summon[s]* Dasein to its ownmost potentiality-of-being-a-self" (Heidegger 1953:269). Like Kant's, this summons is adverbial rather than action- or object-focused, but it is more radically adverbial than Kant's. We cannot "define *what is spoken*" in this call of conscience, since "the call does not say anything . . . 'Nothing' is called to the self . . . , but it is summoned to itself, . . . to its ownmost potentiality-of-being," so "what the call discloses . . . gets interpreted in difference ways in individual Da-sein" (Heidegger 1953:273-4). Heeding conscience "brings the self back from the loud idle chatter of the they's common sense" (Heidegger 1953:296). But conscience cannot replace this chatter with another principle, which would become another chatter; heeding conscience is not a matter of doing the right thing, but of authentically realizing potentialities that are one's *own*.

In a sense, there is nothing "wrong" with bad faith or inauthenticity.²⁷⁷ Sartre criticizes Heidegger for using the term authenticity, claiming that this expression is "dubious and insincere because of [its] implicit moral content" (Sartre 1956: 680). Heidegger too resists "moralizing critique" (Heidegger 1953:167, see too 42-3) and insists that terms like inauthenticity²⁷⁸ "do not express any negative value judgment" but rather merely refer to how "Da-sein is initially and for the most part" (Heidegger 1953:175). Because nothing constrains freedom, existentialists propose no *standards* of authenticity that one "must" or "ought to" live up to. Sartre and Heidegger thus face

²⁷⁷ "[Inauthenticity] would be badly misunderstood if we wanted to attribute to it the meaning of a bad and deplorable ontic quality which could perhaps be removed in the advanced stages of human culture" (Heidegger 1953: H176)

²⁷⁸ Strictly speaking, he is in this passage talking about the term "entanglement," but the same point can be made of "inauthenticity."

Nietzsche's problem of how to develop normativity "beyond good and evil." None of these three moralistically require us to live up to certain standards, but all *invite* us to live in a certain way. In that sense, there is some admonition to authenticity in Heidegger, some reproach of bad faith in Sartre. Heidegger claims, "Da-sein bears witness to a possible authenticity of its existence . . . [and] demands it of itself" and "conscience . . . summons Da-sein to existence, to its ownmost [and thus authentic] potentiality-of-being-a-self" (Heidegger 1953:266-7, 294). And Sartre "bring[s] moral judgment to bear" on one's existential situation:

[T]here is . . . dishonesty if I choose to state that certain values exist prior to me . . . Suppose someone says to me, "What if I want to be dishonest?" I'll answer, "There's no reason for you not to be, but I'm saying that that's what you are, and that the strictly coherent attitude is that of honesty." (Sartre 1993:45)

Those in bad faith deceive themselves about their freedom to define their own values, and even while recognizing that, by virtue of this freedom, bad faith is a possibility for them, Sartre (and Heidegger) insist upon calling their readers to something more, to honest authenticity. But Sartre's reference to "coherence" gets at something deeper about the strategies of Sartre and Heidegger (and Nietzsche and Kierkegaard). By drawing attention to the possibility of authenticity, Sartre and Heidegger show how "the attitude of refusal and flight which remains possible is despite itself the free assumption of what it is fleeing" (Sartre 1956: 680). Reading Sartre and Heidegger makes it impossible to continue in bad faith in the same way one did before.²⁷⁹ In *Sickness unto Death*, Kierkegaard anticipates this approach, taking readers from the unreflective inauthenticity that characterizes everyday life through a process of self-awareness that ends with either "demonic" inauthenticity that "rebel[s] against all existence" (Kierkegaard 1989:104-5) or a breakthrough into authenticity (or "faith"). Similarly, Sartre and Heidegger push their readers towards the recognition of their own freedom, in the context of which even bad faith takes on a more "authentic" character, since it is deliberately chosen in defiance of freedom. In that sense, at least, witnessing the possibility of authenticity is a sort of call to authenticity. Existentialism, by repeatedly bearing this witness, continually issues this call.

²⁷⁹ Or, better, this is impossible while reading (and reflecting on) Sartre and Heidegger. As Sartre, Heidegger, and especially Kierkegaard point out, it is remarkably easy, even immediately after being struck with the angst of one's freedom, to throw oneself back into the distractions of everyday life.

6. Being-with/for-Others.

Given existentialism's emphasis on human freedom from-within, it might seem solipsistic. Being-in-the-world opens space for others in structuring one's situation, but only – apparently – in the same way as mere objects. But, as Kant insists in his moral philosophy and as is evident in our ordinary lives, other human beings are not mere objects, and others structure our existential situation in particularly profound ways. In fact, although I have sidelined it until now, the theme of others (or “the Other”) plays an extremely important role within existentialist philosophy.

Heidegger and Sartre see our primary orientation towards others in radically different ways. For Heidegger, the primary existential orientation towards others is “Mit-sein” or Being-with. Heidegger's discussion of Being-with begins with questioning “what we might call the ‘subject’ of everydayness” (Heidegger 1953:114), that is, who it is that typically is in-the-world in one's ordinary, everyday, being-in-the-world. The answer to this might seem obvious: oneself. But Heidegger argues that, paradoxically, “the who of everyday Da-sein is precisely not I myself” (Heidegger 1953:115). “[A]n isolated I without the others is in the end . . . far from being given initially . . . [T]he ‘others’ are always already there with us in Being-in-the-world” (Heidegger 1953:116). Just as Heidegger insists that who one is is a matter of how one exists in and towards the concrete situation in which one finds oneself, so he adds that one always also finds one's situation to be a situation of being “with” others. Unlike the world of handy objects in which we find ourselves, others are not handy tools for our use, but those *with whom* we use objects. To work with a hammer involves putting the hammer to use for one's purposes; to work with another person involves working alongside that other, sharing purposes with that other, seeing the world through common eyes. “The others’ does not mean everybody else but me—those from whom the I distinguishes itself. They are, rather, those from whom one mostly does not distinguish oneself” (Heidegger 1953:119). In fact, because human being is so oriented by the expectations and perspectives of others, “Being-with existentially determines Da-sein even when an other is not [actually]²⁸⁰ present and perceived” (Heidegger 1953:120). Moreover,

[In] being-with . . . , as everyday being-with-one-another, Da-sein stands in subservience to the others. It itself is not; the others have taken its being away from it. The everyday possibilities of being of Da-

²⁸⁰ Heidegger uses his technical term “factically” here.

sein are at the disposal of the whims of the others. These others are not definite others. On the contrary, any other can represent them. What is decisive is only the inconspicuous domination by others that Da-sein as being-with has already taken over unawares . . . [B]eing-with-one-another as such creates averageness. It is an existential character of the they. (Heidegger 1953:126-7)

“The they,” for Heidegger,²⁸¹ is the “subject” that takes the place of the “I” in Da-sein’s everyday activities. Insofar as human being is being-with-others, I am always open to losing my unique being-towards-possibilities and instead merely thinking and doing what everyone thinks and does. I might engage in various activities because they are activities that “one in my position” should be doing, in a sense letting “them” – that is, the diffuse, widespread, and internalized expectation of those with whom one exists – make existential choices for me.

This “averageness” and “subservience” should sound similar to Sartre’s conception of “bad faith.” “Being-with-one-another,” which is “an essential tendency of Da-sein” that “level[s] down all possibilities of being” into possibilities available to “the they” and thereby “takes the responsibility of Da-sein away from it” (127) is precisely what Heidegger identifies as “inauthenticity.” Angst, in bringing Da-sein face-to-face with its being-toward-death and thereby its own, unique possibilities, “takes away from Da-sein the possibility of understanding itself, falling prey, in terms of . . . the public way of being interpreted. It throws Da-sein back . . . upon its authentic potentiality-for-being-in-the-world” (Heidegger 1953:187). The initial and dominant mode of being is an inauthentic being-with-others that loses its ownmost possibilities through interpreting itself in the light of public norms. Authenticity achieves independence from a condition in which one initially and for the most part finds oneself, subservient to expectations of others in the form of “the they” (but cf. Heidegger 1953:298 for an account of authentic being with others).

Like Heidegger, Sartre recognizes that one does not see others simply as things-in-the-world defined by one’s own projects, but Sartre argues that our primary orientation towards others takes the form of “being-*for*-others” rather than “being-with-others,” and he elucidates this being-*for* with an example:

Let us imagine that moved by jealousy, curiosity, or vice I have just glued my ear to the door and looked through a keyhole . . . The door

²⁸¹ The German here is “das man,” which refers to the impersonal pronoun, as in “one [*man*] would do such-and-such”.

and the keyhole are both instruments and obstacles; they are presented as 'to be handled with care'; the keyhole is given as 'to be looked through close by and a little to one side,' etc. . . . [T]his ensemble exists only in relation to a free project of my possibilities. Jealousy, as the possibility which I am, organizes this instrumental complex by transcending it . . . But I am this jealousy; I do not know it . . . This ensemble . . . we shall call situation. This situation reflects to me at once both my facticity and my freedom; on the occasion of a certain objective structure of the world which surrounds me, it refers my freedom to me in the form of tasks to be freely done . . . I cannot even define myself as truly being in the process of listening at doors. I escape this . . . by means of my transcendence [i.e., freedom]. (Sartre 1956: 347-8)

So far, nothing in this situation involves the recognition of another as an other. No doubt, I am spying on some other person through the keyhole, but this person is merely an object of my activity; the meaning of this other is given by my jealousy. In my being-in-the-world, my free possibilities are possibilities in the context of a concrete situation, but this situation, too, has its meaning only by virtue of my freedom (manifested here as jealousy).

But all of a sudden I hear footsteps in the hall. Someone is looking at me! What does this mean? It means that I am suddenly affected in my being and that essential modifications appear in my structure . . . I now exist as myself . . . I see myself because somebody sees me . . . my freedom escapes me in order to become a given object . . . The Other has to make my being-for-him be insofar as he has to be his being. Thus each of my free conducts engages me in a new environment where the very stuff of my being is the unpredictable freedom of another. Yet by my shame I claim as mine that freedom of another. (Sartre 1956: 349-51)

In feeling ashamed at the look of another – or, as Sartre usually puts it, “the Other” – I “am” for the other. And for Sartre, this is not merely a matter of being seen as an object, since it is precisely my freedom (or “transcendence”) that, in my shame, I feel as being objectified by the other: “my transcendence becomes for whoever makes himself a witness of it . . . a purely established transcendence, a given-transcendence” (Sartre 1956: 352). Here “the Other [is] given to me directly as a subject although in connection with me” (Sartre 1956: 341). As in the case of Heideggerian being-with, being-for-others need not even depend upon the presence of actual others. Even if I recognize that the “footsteps” were a “false alarm,” I can continue to feel ashamed and may even “give up the

enterprise” in the light of my new awareness of myself as being-for-the-Other, not the “concrete historical event” of a particular other person appearing but the always present possibility of being-seen, the constant “relation to every living person” (Sartre 1956: 370, 373).

For Sartre, this objectivity-before-an-Other can be a source of bad faith. Insofar as I merely accept the Other’s judgment of me, relinquishing my being-for-itself into a being-for-another, I fail to acknowledge that “I am [in my freedom] not what I am [for-others].” But for Sartre, being-for-others is a “fact” (Sartre 1956: 377), a “contingent necessity” (Sartre 1956: 337), something that we cannot ignore but that threatens the very freedom at the core of our existence. There is an ongoing struggle with others in which one seeks to avoid being objectified through objectifying them in turn. The one looking through the keyhole can turn back on the one coming up the stairs, looking at the Other and thereby making the Other into an object. Instead of losing one’s objectivity to the look, one can overcome one’s shame with a pride that says, in effect, “What are you looking at?” The contingent necessity of being-for-the-Other puts each person in a constant tension between “being-looked-at” and “being-looking-at” (Sartre 1956: 373). In fact, “I am responsible for the very existence of the Other” (Sartre 1956: 382), or better, “the Other and I are . . . co-reponsible for each Other’s existence” (Sartre 1956: 383), since it is only by my freely taken up projects that I can feel ashamed at the look and thereby give the look meaning as a look, but in giving the look meaning as a look, I relegate its meaning to the Other. “One must either transcend the Other or allow oneself to be transcended by him. The essence of the relations between consciousnesses is not being-with; it is conflict” (Sartre 1956: 555). Or, more concisely, “hell is other people” (Sartre 1989:45).

For both Heidegger and Sartre, others (or the Other, or “the they”) threaten authentic existence. Heidegger sees this threat as intrinsic to Da-sein’s being-in-the-world, while Sartre emphasizes being-for-others as something distinctive from and in conflict with being-in-the-world as such. More importantly, we might say that while Heidegger’s conception of being-with involves “a we-subject” (Sartre 1956: 551), Sartre’s involves an awareness of oneself as a “you,” an object of another’s attention.²⁸² But both see others as a threat to existential authenticity.

²⁸² We might think, in fact, of both Heidegger and Sartre as drawing on the experience of others in order to move beyond the straightforward dichotomy between transcendental and empirical perspectives on human being. The transcendental perspective, in Kant, is a sort of first-person perspective, the from-within perspective of an “I” on the world, while the empirical is a sort of third-person perspective from-without on things in the world (even on oneself as a mere object in the world). Heidegger, in his conception of being-

7. Heteronomous Existentialism

Existentialism continues to play important roles in contemporary conceptions of human being, but in recent years, a new approach has shifted away from Sartre's self-centered existentialism. Shortly after Sartre published "Existentialism is a Humanism" (1946, cf. Sartre 1993), which became the classic statement of modern existentialism, Heidegger rejected the label "existentialist" and distanced himself from Sartre in his own "Letter on Humanism" (1947). In that work, Heidegger reminds his readers of his initial emphasis in *Being and Time* on "the explicit and lucid formulation of the question of the meaning of being" (Heidegger 1953:7). For most of *Being and Time*, Heidegger focuses on Da-sein,²⁸³ or distinctively human being; and Heidegger's account of human being proves conducive for the sort of existential philosophy in which Sartre is particularly interested. But in this later work, Heidegger recalls his interest in the question of Being itself, reformulates his conception of Da-sein in terms of "Being," and clarifies that his conception of "possibility" (and even "freedom") has a different, less self-oriented tone than Sartre's.

In his "Letter on Humanism" and other late writings, Heidegger emphasizes the importance of "thinking" as a form of "letting" oneself "be claimed by Being" (Heidegger 1977: 218, cf. 223). Heidegger rejects the Sartrean view that "the essential worth of man . . . consists in his being the substance of beings as the 'Subject' among them" (Heidegger 1977: 234, cf., e.g., Sartre 1993:50). While admitting, with Sartre, that "by the assessment of something as a value what is valued is admitted only as an object for man's estimation," Heidegger takes this to be a reason to reject values, or at least to "think against values": "Every valuing, even where it values positively, is a subjectivizing. It does not let beings: be. Rather, valuing lets beings: be valid—solely as the objects of its doing" (Heidegger 1977: 251). Heidegger shifts from human freedom as the "foundation of all essences" (Sartre 1956: 567) to freedom that "conserves its essence" in "the realm of the truth of Being" (Heidegger 1977: 247), from "man [as] lord of beings" to "Man [as] the shepherd of Being" (Heidegger 1977: 245). Even while Sartre emphasizes situation, he

with, broadens the notion of subjectivity to include a first person plural perspective on the world, the perspective that "we" take on the world, a perspective in which my own, unique, from-within perspective can get lost. Sartre, instead, emphasizes what we might think of a second-person perspective, a perspective not-my-own from which I am addressed, a perspective on me that sees me, not strictly as an object, but as a "transcended transcendence," as something free that is made into an object through being addressed.

²⁸³ He justifies this focus on the grounds that "a prior suitable explication of a being (Da-sein) with regard to its being" is required for formulating the question of Being (BT 7).

fundamentally locates the meaning of situation in humans' free responses, but Heidegger insists,

Man is . . . “thrown” from Being itself into the truth of Being . . . in order that beings might appear in the light of Being as the beings they are. Man does not decide whether and how beings appear, whether and how God and the gods or history and nature come forward into the clearing of Being . . . The advent of beings lies in the destiny of Being. But for man it is ever a question of finding what is fitting in his essence that corresponds to such destiny; for in accord with this destiny man as *ek-sisting*²⁸⁴ has to guard the truth of Being. Man is the shepherd of Being. (Heidegger 1977: 234)

In this re-emphasis on human freedom, human being is conceived as something that *responds* to Being, that “lets” Being claim it, that exists “into the openness of Being” (Heidegger 1977: 252).

At the end of his “Letter on Humanism,” Heidegger draws on the example of Heraclitus, who, engaged in the “everyday and unexciting occurrence” of “warming himself at a stove,” nonetheless presents a model of thinking that lets itself be claimed by Being. When Heraclitus says to his visitors, “Here too the gods come to presence” (Heidegger 1977: 257), he invites them to think, to let themselves be claimed, to abide in the place where Being – the god(s) – presences itself. Like more traditional, Sartrean (and, arguably, early Heideggerian) existentialism, the notion of “thinking” emphasized by the late Heidegger is a way of being human that is unconstrained by particular rules or forms of thought (being “against logic” (Heidegger 1977: 249)). “Thinking,” like authenticity, is adverbial, a matter of *thinkingly* engaging one’s world. But unlike Sartrean and early Heideggerian emphases on authenticity as a recognition of one’s freedom and ownmost possibilities, thinking involves being receptive before Being. Here “be-ing” is a way of “be-ing claimed” rather than an active “be-ing” as projecting. Although Heidegger seeks to move beyond (or prior to) categories like subjectivity—objectivity or autonomy—heteronomy, we can think of this effort to de-center human being as a sort of existential heteronomy. Where Sartre sought an absolute freedom of activity that takes responsibility for itself and thereby creates a world, Heidegger seeks a thinking that is fundamentally receptive to the call of Being.

This emphasis on what I am calling heteronomous existentialism did not end with Heidegger, and his ontological formulation of it has not

²⁸⁴ Heidegger’s neologism, drawing attention to the root meaning of “exist” as “to stand out (from),” from the roots *ex* (out) and *sistĕre* (to stand).

turned out to be the most important contemporary response to subject-centered, Sartrean existentialism. For that more radical response, one that not only rejects the emphasis on subjectivity but rejects the “ontological” tradition of which both Heidegger and Sartre are a part, we must turn to the work of Emmanuel Levinas.

Levinas continues Heidegger’s interest in moving beyond concrete, determinate “valuings” or “understandings” of beings in order to “let it say something to us” (Heidegger 1977:442). For Levinas, however, any “letting” will fail as long as one remains focused on Being:

In subordinating every relation with existents to the relation with Being the Heideggerian ontology affirms the primacy of freedom over ethics. To be sure, the freedom involved in the essence of truth is not for Heidegger a principle of free will. Freedom comes from an obedience to Being; it is not man who possesses freedom; it is freedom that possesses man. But the dialectic which thus reconciles freedom and obedience in the concept of truth presupposes the primacy of the Same, which marks the direction of and defines the whole of Western philosophy. (Levinas 1969: 45)

The “primacy of the Same” refers to a general goal of Western metaphysics to overcome irreducible otherness (or “alterity”). Even as Heidegger moves beyond traditional, Western-metaphysical categories, he remains part of a continuous tradition in philosophy that seeks to “apprehend the individual not in its individuality but in its generality” (Levinas 1969: 44), since Heidegger, too, “subordinates the relationship to someone, who is an existent . . . to a relation with the Being of existents, which, impersonal, permits the apprehension, the domination of existents” (Levinas 1969: 45). Put another way, Heidegger’s remains a philosophy of “knowing” rather than an “ethics.”²⁸⁵

Levinas turns from “ontology” – the problem of Being – to “ethics.” But Levinasian ethics is not Kantian, not an effort to discern universal moral principles by which all ought to conduct themselves. Instead, Levinasian ethics is the

calling into question of my spontaneity in the presence of the Other . . . The strangeness of the Other, his irreducibility to the I, to my

²⁸⁵ In this section, I will not pursue the question of how fair this interpretation is to Heidegger. Arguably, Heideggerian ontology is also “ethical” in Levinas’s sense; Heidegger does not see the Being of beings as a universal under which individual beings can be subsumed, nor is Heideggerian “thinking” equivalent to “knowing” in the way that Levinas suggests.

thoughts and my possessions, is precisely accomplished as a calling into question of my spontaneity. (Levinas 1969:43)

As in later Heidegger, Levinas de-emphasizes subjectivity and spontaneity in favor of a privileged heteronomy and “total passivity” (Levinas 1969:88, 1997:87, 1998:110). But for Levinas, this heteronomy makes sense only in terms of responsibility for a concrete Other. This “passivity more passive . . . than any passivity” is the “passivity of the ‘for-another,’” a passivity “in which no reference, positive or negative, to a prior will enters” ((Levinas 1998a:50-1). The “Other” here is not “Being,” but rather the “face” of the “neighbor” (Levinas 1997: 92, Levinas 1969:194f.) Levinasian “ethics” thus cannot be a matter of laying out fundamental moral principles, but only a recurring provocation to truly see the infinite Otherness of this face, a clearing away of distracting “categories” that prevent such seeing.

Levinas’s overriding interest throughout his “ethics” is to highlight the originary status of the concrete other. His general strategy can be understood as a revaluation of Sartre’s account of the Look, such that Levinas prioritizes being-for-another over being-for-itself. As in the case of Sartre, being-for-another is provoked through the encounter with a concrete other. Something “sensible” and “still graspable, turns into total resistance to the grasp” (Levinas 1969: 197). Just as, for Sartre, footsteps heard in the hall become an Other that transforms one’s being into being-for-another; the Levinasian encounter with the Other begins with something sensible and graspable, but quickly transforms “by the opening of a new dimension. For the resistance to grasp is not produced as an insurmountable resistance, like the hardness of a rock against which the effort of the hand comes to naught, like the remoteness of a star in the immensity of space. The expression . . . introduced into the world does not defy the feebleness of my powers, but my power of power²⁸⁶” (Levinas 1969: 197). This sensible expression defies the very power by which I am able to grasp anything at all, to make the world “my own.”

For Levinas, the sensible expression that opens this new dimension is “the face.” What is perceived²⁸⁷ in the face of the other is

²⁸⁶ I’ve modified Lingis’s translation of “mon pouvoir de pouvoir” here.

²⁸⁷ Technically, Levinas uses the term “appresentation” (Levinas 1997:166) rather than perception. Levinas borrows the concept “appresentation” from Husserl, who used it to describe how objects are presented to consciousness. Only a single facet of an object is directly presented – one sees it only from one angle – but the whole object is perceived because its not-directly-seen aspects are “appresented.” Levinas appropriates and importantly modifies this notion, but I pass over these subtleties in my discussion.

not something, but “unicity and alterity,” the wholly-unique and wholly-other.

The face resists possession, resists my powers . . . The face, still a thing among things, breaks through the form that nevertheless delimits it. This means concretely, the face speaks to me and thereby invites me to a relation incommensurate with a power exercised, be it enjoyment or knowledge. And yet this new dimension opens in the sensible appearance of the face . . . The face at the limit of holiness . . . is thus still in a sense exposed to powers. (Levinas 1969: 197-8)

When another speaks, she ceases to be merely an object to categorize in terms of my projects and possibilities. Instead, I am addressed *by* the other and thereby made responsible for the other. Neither I nor the other are “objects” in the traditional sense. The other is a subject by and for whom I can be held to account, and “I” am one to whom speech is addressed.²⁸⁸ The emphasis on language is not tied to particular words spoken (“the Said”), but simply to the fact that by “speaking to me” (“Saying”), the face reveals itself as something that is not always-already

²⁸⁸ While both the distinctive vulnerability characteristic of the human face and the role of language play important roles in Levinas’s account of the Other, neither “speaking” nor the “face” need be understood strictly literally. Levinas recounts how “the backs of others” waiting in line for news of imprisoned relatives can be “the face” insofar as these reveal “the extreme precariousness” of the other and thereby “speak” to us (Levinas 1997: 167). Levinas’s basic point here is similar to that of Sartre’s account of the Look; in the face of the Other, one is addressed: “It is no longer a question of the Ego [the I] but of me” (Levinas 1997: 120). In contrast to Kant’s dichotomy between the first-personal “I” of transcendental anthropology and the third-personal “that” of empirical science, Levinas privileges the second-person perspective of another for whom one is a “me.”

That said, Levinas occasionally resists second-personal language, specifically contrasting his account with any I-thou relationship and coining the neologism “Illeity” (from the French “Il” (he)) to describe what he has in mind. But when explaining what he means by illeity, he says,

The illeity in the beyond-being is the fact that its coming toward me is a departure which lets me accomplish a movement toward the neighbor. The positivity of this departure, that which makes this departure . . . be more than a term of negative theology, is my responsibility for the others. Or, one may say, it is the fact that the others show themselves in their face. There is a paradox in responsibility, in that I am obliged without obligation having begun in me. (Levinas 1997: 119).

The point of illeity is that there is an *il*, a *he*, that addresses me and thereby imposes responsibility upon me. The shift away from I-thou is really towards a him-me language, where the “me” is doing most of the work. But the perspective of “me” as the object of address of another is really the perspective of seeing myself as a “you” that is addressed by another, a “you” that is responsible to the other. In that sense, Levinas’s illeity is, misleadingly, an allusion to a second-personal standpoint.

defined by my knowledge of or desire for it; by speaking, the face shows itself to be an Other capable of addressing me.²⁸⁹

Importantly, for Levinas, the Other who addresses me is neither merely a concrete other with determinate qualities one can recognize and accommodate nor a universal “humanity” that requires respect. Levinas’s Other outstrips conceptualization. The very attempt to “understand” the other is a way of trying to reduce the Other to the Same, and thereby to make oneself, one’s own subjectivity, the center of the universe.

For Levinas, encountering the Other is originary. One comes to be at all only through responsibility to the absolutely Other: “[T]he identity of the subject comes from the impossibility of escaping responsibility” (Levinas 1997:120). “Responsibility preced[es] freedom” (Levinas 1997:94, see too 119, 121) and thus “cannot have begun in my commitment, in my decision” (Levinas 1997:117). Against both Sartre and Kant, Levinas argues,

Existence is not in reality condemned to freedom, but it is invested as freedom . . . To philosophize is to trace freedom back to what lies before it . . . Knowledge as a critique, as a tracing back to what precedes freedom, can arise only in a being that has an origin prior to its origin—that is created . . . The presence of the Other, a privileged heteronomy, does not clash with freedom but invests it. The shame for oneself, the presence of and desire for the other are not the negation of knowing: knowing is their very articulation. The essence of reason consists not in securing for man a foundation and powers, but in calling him into question and in inviting him to justice. (Levinas 1969: 84-5, 88)

It is only in response to the responsibility felt in the presence of another that one can make sense of one’s own freedom. Moreover, as his reference to “knowing” makes clear, Levinas sees a crucial epistemic role for the Other as well. *All* knowing is merely an articulated response to an encounter with the face of an Other.

Even while infinitely exceeding my powers, however, the face is also “in a sense exposed to my powers” (Levinas 1969:198). Just as, for Sartre, one can overcome being-for-another through subjugating the

²⁸⁹ Levinas makes much of the importance of language in this respect, and develops an account of the distinction between the Saying, the irreducible address of the Face, and the Said, the particular and comprehensible contents of any particular Saying. Metaphysics (and Western philosophy) are then described (and criticized) as attempts to reduce the Saying to the Said, the Face to the categories in terms of which that face can be something “for us.”

other to one's freedom, so, for Levinas, one can "kill" the Other. To kill the Other "is not to dominate but to annihilate" (Levinas 1969: 198). Every other that is not wholly Other can be dominated, and even cutting off the physical life of an organism is really a way of dominating, rather than annihilating, that organism. By putting it to my purposes – as food, as a vanquished threat, or even just as sport – I dominate an organism by "killing" it in the literal sense. But to kill the Other, in the Levinasian sense, is to "totally negate" the Other, to reject its hold over me, its infinite incomprehensibility. For Levinas, "The Other is the sole being I can wish to kill" (Levinas 1969: 198), precisely because every other being can, in some sense or another, be made to serve my purposes. But by calling me into question, by imposing an infinite responsibility on me through its address, the Other challenges me to assert myself against it, to impose the order of the Same. In that sense, one "kills" the Other whenever one interprets the Other in terms of familiar categories, as just an object of one's projects. "Killing" is a Levinasian equivalent of "bad faith," where one pretends that the infinite otherness of the Other can be subsumed into familiar categories. To take an extreme but salient example, when one sees another merely as an instance of a "humanity" that universal morality requires one to respect, one thereby kills the Other, even if, out of "respect," one promotes the happiness and well-being of what has become the object of one's attention.²⁹⁰

Like Sartre,²⁹¹ Levinas maintains that in the face (or look) of the Other, "one must either transcend the Other or allow oneself to be transcended by him" (Sartre 1956: 555). But whereas Sartre takes this to imply that "The essence of the relations between consciousnesses is . . . conflict." (Sartre 1956: 555), Levinas sees the essence of consciousness itself in the responsibility for the Other that tempts to murder but also prohibits it. Precisely because of the vulnerability of the Other to being "killed," the speech of the other, whatever its particular content, is always also "the primordial expression, the first word: 'you shall not commit murder'" (Levinas 1969: 199). The face imposes an "ethical resistance," a sense of responsibility:

The thought that is awake to the face of the other human is not a thought of . . . ²⁹², a representation, but straightaway a thought for . . .

²⁹⁰ Here there is a close parallel with the bad faith involved in obeying the moral law as a "categorical" imperative.

²⁹¹ But against Heidegger: the following quotation continues "The essence of the relations between consciousnesses is not *Mitsein*."

²⁹² Ellipses original.

. , a nonindifference for the other, upsetting the equilibrium of the steady and impassive soul of pure knowledge. (Levinas 1997: 166)

The Other is not an other of whom we are aware, but the Other for whom we are responsible, precisely insofar as the Other is both “absolutely other” and “vulnerable” (Levinas 1969:199; 1997:17, 121, 1998b:29). By imposing absolute responsibility, “The ethical relationship . . . puts the I in question” (Levinas 1969: 195, cf. Levinas 1997:16). The I becomes a “me,” passive before the Other who addresses me. In this relation, “the Same welcome[s] the Other, not by giving the Other to itself as a theme (that is to say, as a being) but by putting itself in question,” which occurs “precisely when the Other has nothing in common with me, when the Other is wholly other, that is to say, a human Other . . . [, w]hen, through the nakedness and destitution of his defenseless eyes, he forbids murder and paralyzes my impetuous freedom” (Levinas 1997: 16). In this way, the primordial “letting” that Heidegger sought in Being is found, for Levinas, in the overcoming of the Same, of the I, of spontaneity, through one’s responsibility for another.

Moreover, for Levinas, the absolutely free Sartrean subjectivity that rises above given codes or laws *proceeds from* a heteronomy that is responsible to the absolutely Other. Precisely because the Other is infinitely Other, no set codes or principles can be sufficient for guiding one’s response to an Other. Absolute freedom is necessary because no fixed laws can guarantee that one will avoid cruelty, that is, that one will fulfill one’s responsibilities to the Other (see Levinas 1997:23). This emphasis on the impossibility of subsuming responsibility under fixed categories of reason marks a sharp contrast from the Kantian attempt to derive a “formula” of morality from the conditions of possibility of moral responsibility as such. Levinas’s rejection of such a formula follows from his thinking about the radical Otherness of the Other, and as in the case of Sartre and Heidegger’s rejection of similar formula, responding to the Other is not a matter of doing any particular thing, but of having a certain attitude, of doing whatever one does with a sense of responsibility before something infinitely other to and for whom one is infinitely responsible.

8. Kantian Existentialism.

Existentialism raises three central challenges to Kant’s conception of the human being. First, in its emphasis on Being-in-the-world, existentialism calls into question Kant’s privileging of “objective”

experience and his related endorsement of a limited scientific realism. In this context, existentialism also challenges Kant's sharp distinctions between cognition, feeling, and volition. Second, with its emphasis on absolute freedom (or absolute Otherness), existentialism challenges the ways in which Kant seeks to elucidate individual freedom and our relations with others in terms of the moral law. While Kant insists that the moral law provides the law of our freedom and a structure of respect for others, existentialists insist that freedom (and/or the Other) necessarily outstrip *any* law.²⁹³ Finally, the emphasis on absolute freedom (or infinite Otherness) brings with it an ideal of authenticity. Where Kant saw only the good will as good without qualification, existentialists call us to a way of being within which the submission of one's freedom to universalizable laws appears as an inauthentic enslavement to "the they," a form of "bad faith" in which one pretends that some law arises "necessarily," or a way of trying to ignore the infinity of the Other by relating to that Other in accordance with general norms. This section shows how each existentialist challenge can be answered.

Kantian Being-in-the-world.

With respect to Being-in-the-world, existentialists show that the account of empirical knowledge offered in Kant's *Critique of Pure Reason* is an account of only a very special sort of "objective" knowledge. Most people, most of the time, in their "average everydayness," do not experience the world as merely a succession of alterations amongst objects in Euclidean space. Objects are understood in terms of practical purposes and projects, and even their basic spatial and temporal attributes are defined in terms of goals we pursue and tools at our disposal. Kant should not, and does not have to, deny any of these existentialist claims. To take one mundane example, Kant points out in his *Anthropology* that "German miles (which are not . . . indicated with milestones . . .) always become shorter the nearer we are to a capital (e. g., Berlin), and longer the farther we are from one (in Pomerania)" (7:234).²⁹⁴ Kant's epistemological focus on objective spatial and temporal

²⁹³ Moreover, given the collapse of the dichotomy between cognition and volition, this freedom has implications for our knowledge of the world as well. These implications are particularly pointed in the case of heteronomous existentialists, who emphasize that the Other (or, for Heidegger, the Being of beings) always outstrips any determinate categories with which one tries to make sense of it. While Kant seeks to subsume all knowledge into a coherent set of a priori categories, (heteronomous) existentialists see this as an ultimately misguided privileging of "reason" or "the Same" over the infinite "Other."

²⁹⁴ Kant uses this spatial example as an analogy for an even more interesting temporal one, "the phenomenon that a human being who has tortured himself with boredom for the greatest part of his life, so

relations does not preclude his awareness of and attention to the *non*-“objective,” ordinary, lived experiences of human beings.

But where Kant and existentialists differ regarding the priority of the “life-world” and the objective, scientific world. Heidegger emphasizes that any object of our experience “is always already a thing at hand in the surrounding world and precisely not ‘initially’ merely objectively present ‘world-stuff’” (Heidegger 1953:85, see too BT H 106), while Kant insists that, at least in some sense, an objective world presents itself to cognition initially. This “initially,” however, is neither *temporal* nor *phenomenological*. Kant does not claim – at least not in his transcendental anthropology²⁹⁵ – that we *first* have cognitions and *then* practical purposes.

Even the first *Critique* already isolates an artificially abstract form of “experience” – objective, empirical knowledge – Kant recognizes that such experience always already involves *both* concepts and intuitions. One can abstract concepts away from the experiences within which they are typically situated, but, from-within, one does first find oneself with concepts or intuitions and then put them together to get a coherent experience. What Kant argues, however, is that we can make sense of the nature of these always-already-integrated objective experiences only by seeing that they involve two separate components that are synthesized into a single cognition.²⁹⁶ One might make a similar move in relating the “experience” that Kant analyzes in the *Critique of Pure Reason* to the everyday experience existentialists emphasize. Kant need not claim that most or even any of our lived experience is reducible to merely objective knowledge disconnected from feelings and purposes. And Kant can certainly admit that our cognitions of objects employ purposive and affective concepts (such as “hammer” or “too heavy” or “fearsome”). Rather, what Kant claims is that in any cognition, whether or not that cognition is related to feelings and volitions, there is an element that is

that every day seemed long to him, nevertheless complains at the end of his life about the *brevity* of life” (7: 234).

²⁹⁵ Within his empirical psychology, Kant does claim that cognitions precede feelings and volitions, but these cognitions need not be “objective” in the sense relevant here. Moreover, while there are, for Kant, some cognitions that are not conjoined with feelings or desires, many (perhaps most) cognitions are part of a causal series that culminates in a volition, and in these cases, one’s lived experience may well be primarily of the cognition-feeling-volition complex; only for experimental psychologists dissecting human mental states will the separation and causal relations emerge clearly.

²⁹⁶ The details of Kant’s account here are scrutinized by Heidegger in his *Kant and the Problem of Metaphysics* (Heidegger 1997), where Heidegger appeals to Kant’s allusion to a “common root” of sensibility and intuition to argue for a primordial unity between the two. For one important Kantian response to this particular argument, see Henrich 1955.

directed towards objects as such (independent of their role as tools for my purposes or sources of pleasure):

there is always a great difference between representations belonging to cognition, insofar as they are related merely to the object and the unity of the consciousness of it, and their objective relation where, considered as at the same time the cause of the reality of this object, they are assigned to the faculty of desire, and, finally, their relation merely to the subject, where they are . . . [related] to the feeling of pleasure. (20:206)

Even if most of our relations with objects include all three components – mere cognition, desire, and pleasure or pain – there is still, for Kant, a distinction between these components of lived experience.

As a first step towards defending this claim, we might note that modern science, and even ordinary “objective” judging, at least *seems* to fit Kant’s description of “cognition” as such. Even if some aspects of science (especially engineering and technological applications) are directed specifically towards practical purposes, and even if a great deal of scientific research has practical ambitions in view, some areas of science are particularly praised as “pure science” for seeking merely to adequately describe and explain our world.²⁹⁷ Superstring theory and accounts of the origin of the universe may come to have practical applications, but they are primarily ways simply of conforming our ways of thinking to a world that exists independent of our purposes. Even such things as evolutionary biology and modern chemistry are defended first and foremost as accurate representations of the world, not practical tools for medicine or industry. Many important debates within science (such as Galileo’s insistence that Copernican astronomy is not merely a good *model*, but *true*) and even the ambition for scientific progress itself seem best explained in terms of scientists’ striving to accurately map their beliefs onto the world. Relatedly, scientific knowledge – and “objective” knowledge more generally – at least *seems* less subjective, less tied to the particular situations or tendencies of individual knowers, than the average, everyday experience on which Heidegger and Sartre focus. In that sense, objectivity seems to get at something important about the world, something about how the world is, independent of our particular

²⁹⁷ We might even put this in terms of a sort of experiment: Granted that Kant can allow that cognition, volition, and feeling are often and perhaps usually connected, how can we decide whether they are in fact distinguishable, as Kant claims, or whether they are indistinguishable, as the existentialists claim? If Kant is correct, but not if the existentialists are correct, then we would expect cases in which we find cognitions without relation to our desires or practical purposes. And it looks like we do find this in many theoretical sciences.

relations to it. Finally, some sort of objective awareness of objects seems to be implicated even in our ordinary lived experience. It seems right to say that it is only because hammers are solid, etc. – and because, in some sense, I cognize them this way – that they can be handy for me when I want to pound a nail.

We can put these points in terms of Sartre's example of immersing one's hand in water. It is true that in *some* sense I can "give the name subjectivity to the objectivity which I have not chosen" (Sartre 1956: 412). And it is even true that the *felt* temperature of the water, rather than the temperature on a thermometer, is often our primary engagement with water. (I don't typically need a thermometer to tell me that the water with which I am washing my hands is too hot.) But for all that, it is not arbitrary to describe the relatively high temperature on the submerged thermometer as "objective" and the perception of the water as cold by my hot hand as "subjective." The former, but not the latter, remains constant as long as the condition of the water remains constant, while the latter is subject to wild fluctuations based on my present condition. Partly for this reason, the former, but not the latter, is something that I can share with others. And an explanation of why water that is "really" warm "seems" cold to my hand can be much more straightforward than an attempt to explain why water that is "really" cold "seems" warm to the thermometer.

This defense is not yet sufficient, but it puts us on the right track towards a sufficient response. We can get a sense both for what this defense is missing, and what is right about it, by comparing the optimistic account of science here with Heidegger's own account of how "something at hand with which we have to do . . . turns into something 'about which' the statement that points it out is made" (Heidegger 1953:147). Heidegger rejects the picture of science as merely "keeping our distance from handling" or "merely looking at beings . . . [by] *abstain[ing]* from any kind of use," arguing instead that one must come to "look... *at* the thing encountered in a 'new' way, as something objectively present" (Heidegger 1953:357, 361). This process involves two important components. First, an object that initially is handy and thus inconspicuous becomes "unusable" and thus "conspicuous" (Heidegger 1953:73). When the hammer that we are using is "too heavy" or when a leg on the table on which we are working breaks, we come to see "the character of objective presence in what is at hand," that is, we come to see the hammer or table as something that "just lies there," something that, even when it is handy, "is always also objectively present with this or that appearance" (Heidegger 1953:73). Second, the sort of conspicuous attention provoked by the unusability of things at hand can

itself become a privileged activity; this happens when we begin to make “statements” about such things, and especially when we seek a sort of *uniformity* of judging about the thing.

[B]y and for this way of looking what is at hand is veiled as something at hand . . . [and] forced back to the uniform level of what is merely objectively present . . . This leveling down of the primordial ‘as’ of circumspect interpretation to the as of the determination of objective presence is the specialty of the statement. Only in this way does it gain the possibility of a pointing something out in a way that we can sheerly look at it. (Heidegger 1953:158)

The reference here to a “leveling down” is a deliberate reference to Heidegger’s view of inauthenticity, “being-with-one-another as . . . *averageness*,” the “existential character of the they,” which Heidegger identifies with “the *leveling down* of all possibilities of being” that “takes the responsibility of Da-sein away from it” (Heidegger 1953:127).

For Kant, this analysis of the origin of science captures an important distinction between scientific objectivity and ordinary engagement with handy objects, one that gives good reasons to *privilege* scientific knowing. In particular, “leveling down” of possibilities towards “uniformity” is another way of saying that the scientific project is that form of human engagement with the world in which we seek to develop an understanding that can in principle be the same for all others, regardless of contingent circumstances or particular practical goals. This striving for uniformity explains why the thermometer seemingly gives a more accurate measure of temperature than the hand; it gives a measure that is more *uniform*. By analogy with Kant’s practical philosophy, we might see science itself as a sort of search for “categorical imperatives” within the realm of cognition. Rather than acting on universalizable maxims, science seeks universalizable cognitions. As with the categorical imperative, we “veil” particular projects in order to construct “uniform” judgments. In so doing, we “level down” the world but also thereby *create* a world we can share. And Heidegger rightly draws attention to the connection between *this* sort of knowing and “statements,” or, in Kantian terminology, judgments. Kant models cognition on *judgment*, and both Kant and Heidegger see that emphasizing cognitions we can *say* brings with it an emphasis on cognitions we can *share* and thereby a sort of uniformity that is particularly conducive “objectivity.”

One might, at this point, simply ascribe the difference between Kant and Heidegger to a difference in emphasis. Kant offers transcendental analyses of conditions of possibility of “knowledge” of

conspicuous objects; Heidegger is interested in existential analyses that privilege lived experience of objects already taken up into practical activity. If this were the only difference, Kantians might appropriately supplement their account of knowledge with a Heideggerian account of Da-sein, perhaps emphasizing the advantages of Kant for making sense of the sort of objective awareness that is particularly important in modern science.

But there are two deeper differences between Kant and Heidegger that are more difficult to accommodate. First, Heidegger links scientific objectivity with inauthenticity in part as a way of calling objectivity into question, or, better, bearing witness to alternative possible ways of knowing and thereby “demanding” or “summoning” his readers to these (see Heidegger 1953:266-7, 294). Privileging scientific knowing is a threat to authentically embracing one’s ownmost possibilities of engagement with the world. For Heidegger, Kant’s emphasis is not only not to be preferred; it is positively dangerous in its potential to further entrench the growing deference to scientific objectivity over authentic understanding. Because of the connection of this difference with “authenticity” more generally, I reserve discussion of this point until the end of this section.

The second fundamental difference between Kant and Heidegger relates to which sort of analysis – or better, which way of knowing – is more fundamental: objective cognition or holistic being-in-the-world. For Heidegger, objective cognition is a subsidiary and artificial way of looking at what is “primordially” at hand (Heidegger 1953:158).²⁹⁸ Even for Heidegger, when one seeks objective knowledge, one “sheerly look[s]” (Heidegger 1953:158) constructing the objective world precisely *as* a privileged world which we treat as already there before our projects. For Heidegger, it is constructed in this way by us, for particular reasons, and subsequent to the breakdown of our handy world, so in those senses it is secondary. And it can be constructed only on the basis of veiling the handiness of the handy world, so in that sense its priority is illusory. But Kant and Heidegger agree that at least when we think of the world in terms of objectivity, we explicitly think of it as having a sort of priority.

²⁹⁸ Heidegger even finds oblique confessions in Kant’s own *Critique* of the priority of this primordial and holistic “as.” In his commentary on Kant’s *Critique*, Heidegger makes much of Kant’s allusion to the imagination as a “common root” of concepts and intuitions and of the fact that “thinking and intuiting . . . are not separated from one another like two completely dissimilar things. Rather, as species of representing, both belong to the same genus of pre-presenting in general. Both are modes of the representing of . . .”. (*Kant and the problem of metaphysics*, G148, English 103) Heidegger highlights a supposed common genus of concepts and intuitions, while Kant emphasizes – against Wolff and in contrast to Heidegger – the *irreducibility* of concepts and intuitions to any univocally understood general category of representation.

What Kant shows is how to access this world as a world that exists independent of our practical engagement with it, that exists “objectively.” The result is to show that in our engagement with the world as a world for us to “know,” we employ a priori concepts that justify everyday objective judgments as well as those of math and science, but that are limited in their extent.²⁹⁹

In the end, even if lived human experience does not conform to the sort of “objective” cognition on which Kant focuses, Kant’s account provides transcendental conditions of possibility for at least one important sphere of humans’ being-in-the-world, one connected in particularly important ways with both our efforts to find consensus amongst human beings and our need to provide *justifications* for beliefs about “conspicuous” objects. Moreover, there is at least some reason to think that this sort of objective cognition is a fundamental constitutive element even of the lived experience on which Heidegger focuses (especially insofar as there is a normative dimension to that lived experience). And privileging scientific objectivity has, at the very least, provided human beings with a self-understanding that, as indicated in chapter eight, is already improving human lives.

Freedom and Infinity

With respect to Kant’s conception of freedom, there are three central existentialist critiques of Kant. First, following from the existentialist account of Being-in-the-world, there is an emphasis on freedom as always already *situated*, and thus worldly in a way that Kantian freedom seems not to be. Second, existentialists emphasize an *absolute* freedom at odds with Kant’s attempts to provide universal *laws*

²⁹⁹ To some extent, this difference can be resolved by noting that although Kant and Heidegger both emphasize a from-within perspective on human beings, they have different objects in view. As a phenomenologist, Heidegger emphasizes close and careful *description* of (immediate) *consciousness* of the world. In that sense, Heidegger’s approach is almost a complex form of introspection. But Kant emphasizes *transcendental* analysis of the conditions of possibility of *justifying* our cognition of the ((Levinas 1998a:jective) world. While Heidegger’s anthropological question is directed *towards* the from-within perspective, something like “What does human being-in-the-world look or feel like?”, Kant’s question is directed *from* the from-within perspective, something like “What should I believe (or feel, or do)?” For Kant, but not for Heidegger, normativity is essential to the from-within perspective. And when analyzing the from-within perspective that is *normative*, the perspective from which one is trying to decide what to believe about the world, one’s world is already conspicuous and thus invites the effort to find “objective” measures for belief. In that sense, Heidegger shows that the perspective does not always have the conspicuous character that it does for Kant. But for the from-within perspective from within which one can actually employ philosophy for deliberation about what to think (or do), Kant’s philosophy offers the better account (so far).

of freedom. Third, heteronomous existentialism calls into question Kant's emphasis on "autonomy" and his privileging of a "from-within" standpoint that focuses on the "I" of thought and action rather than the "me" of responsibility (to another).

The first existentialist objection to Kant's account of freedom is that we are thrown into "definite possibilities" (Heidegger 1953:144), so choice is always "choice *in the world*" (Sartre 1956: 617). Existentialists "reject Kant's 'choice of intelligible character'" (Sartre 1956:, 617), juxtaposing their emphasis on situated freedom with Kant's timeless, situation-less free will "independent of all empirical conditions" (5:29, see too A539/B567). In fact, however, as should be clear from chapter one, the existentialist emphasis on freedom-in-situation is remarkably close to Kant. Within his epistemology, Kant shares with the existentialists an insistence that human cognition involves application of concepts to intuitions of *given* objects; human finitude implies that the spontaneity of cognition is always a matter of how we *respond* to a *given* world. And with respect to human *choice*, Kantian freedom comes in our ability to subject our maxims to the moral law. This ability is independent of sensible conditions because (for Kant) the moral law is independent of those conditions. But the maxims that are subjected *to* that law are defined in terms of interests and situations in which we initially and for the most part simply find ourselves. To take Kant's preeminent example, one only has the freedom to refrain from a false promise when one finds oneself in a situation where one is inclined to tell such a promise, a situation such as one involving the need for money one will be unable to repay. The life of freedom is a life of constraining maxims that emerge from one's situation in-the-world in terms of a moral law. What distinguishes free human choice from unfree animal choice is not maxims, but the moral law in terms of which we limit our maxims. And in that sense, there is something timeless about human freedom. But the moral law is always applied to particular contexts, and in that sense Kantian freedom is as situated as Sartrean. The moral dimension of this account is something with which existentialists will take issue, but Kant, like existentialists, takes freedom to be freedom-in-the-world.

The second existentialist challenge to Kant's conception of freedom is the converse of the first. Where existentialists' emphasis on being-in-the-world leads them to reject what they take to be an overly *unconstrained* sense of "intelligible" freedom, their emphasis on the absoluteness of freedom leads them to reject the *constraints* that Kant identifies as intrinsic to freedom. Within the epistemic realm, this means that existentialists reject the absoluteness with which Kant prescribes categories and forms of intuition to the "spontaneous" human

understanding. In Sartre and Heidegger, this involves deprivileging scientific cognition relative to understandings of our lived world. The most extreme form of this rejection of Kantian constraints on cognition comes in earlier, proto-existentialist thinkers like Kierkegaard, who argues that “because the . . . existing knowing spirit is itself in the process of becoming, . . . truth . . . cannot be established absolutely . . . [so] . . . every beginning, when it is *made* . . . is *made* by virtue of a resolution, essentially by virtue of faith” (Kierkegaard 1992:189). For Kierkegaard, this has profound implications since “faith,” for Kierkegaard, is a sort of “passion,” and the more “paradoxical” one’s beliefs, the more passionately one can hold them. Thus Kierkegaard defends a sort of irrationalism, within which the highest knowledge is precisely the faith that breaks through supposed constraints, faith, for instance, that “the eternal . . . has come into existence in time, that God has . . . been born, grown up, etc.” (Kierkegaard 1992:210). Even if they go less far than Kierkegaard’s valorization of paradox, existentialists might point out, especially in the light of the Kuhnian-historicist critiques raised in the previous chapter, that human beings seem precisely *not* to be dependent upon fixed and a priori structures of cognition. We cognize the world in terms that suit our particular projects. Our knowing-the-world is part of a being-in-the-world for which we are responsible, and while this being-in-the-world is always situated, it is never forced – neither by the world itself nor by any “nature” or “transcendental structures” – to interpret that world in any given way. All cognition is interpretation, and interpretation is absolutely free. In this context, Kant’s epistemology looks like veiled essentialism: human knowers “essentially” have spatial and temporal sensibilities of particular kinds, cognitive constraints of various kinds, and so on.

This connection between cognition and projects leads to the more important existentialist critique of Kant: their rejection of Kant’s limitation of the freedom of *choice* by the moral law.³⁰⁰ Sartre puts the point particularly clearly: “Freedom has no essence” (Sartre 1956: 566). To a considerable extent, this claim is connected with the existentialist ideal of authenticity, and I reserve a detailed discussion of the relationship between autonomy and authenticity for the end of this section. But there is also a broader, “metaphysical” point, that Kant is simply wrong to see the moral law as constraining human freedom. With respect to the young student deciding whether or not to join the resistance, Sartre observes that the moral law is unhelpful in actually deciding what to do. More forcefully, one might put the point in terms of

³⁰⁰ Here again, Kierkegaard’s “suspension of the ethical” and Nietzsche’s move “beyond good and evil” pave the way. See especially Kierkegaard 2006 and Nietzsche 1966, 1967.

Kant's own distinction between the freedom that he identifies with "autonomy" and the freedom to which he refers in the *Religion* when he ascribes to human beings a "free choice" of "radical evil." Just as Kant claims that "an incentive [can] have influence on the power of choice of the human being . . . only because this human being incorporates the incentive . . . into his maxim" (6: 24), Sartre insists that "causes and motives have meaning only inside a projected ensemble which . . . is ultimately myself as transcendence" (Sartre 1956: 564). For Kant, the moral law has authority *regardless* of any choice to incorporate this law into a maxim. But, Sartre can ask, why privilege the moral law in this way?

A complete Kantian response to these objections requires rehearsing the arguments in chapter one for the necessity of the transcendental structures of cognition and volition. Put briefly, though, Kant would emphasize that freedom (whether in cognition or volition) has a negative dimension (in that one is free *from* complete determination from sensible influences) but it has this dimension only by virtue of being a positive capacity, a capacity *to* order one's experience, or one's life, into a coherent whole. In the moral case in particular, Kant does *not* think that the incorporation thesis alone is sufficient to ensure that human beings *really* have freedom. Kant rules out freedom being illusory not because we seem free when we make choices, but because our practice of holding ourselves *morally* responsible depends upon a freedom that is *genuinely* undetermined by sensuous incentives. This is compatible with the broadly existentialist point that even if freedom lacks an "essence," one can use choose in ways that fail to properly recognize or live out one's freedom. In that sense, Kant is not far from Sartre's denigration of bad faith, Heidegger's account of inauthenticity, or Kierkegaard's analyses of despair.

A third challenge to Kant's account of freedom comes specifically from *heteronomous* existentialism. Kant divides his anthropology into what we might call a first-person perspective from which one justifies one's thoughts and actions and a third-person perspective that one can take on human beings as objects. Others are conceived either as empirical objects or as practical agents. Kant has no place for the radical Levinasian heteronomy that sees the self as constituted by a response to an inscrutable but vulnerable other. Moreover, Kant leaves no room for practical (or even epistemic) norms or responsibilities that do not in some sense derive from one's own autonomy. Autonomy is central to every aspect of Kant's transcendental anthropology. For Kant, as for Levinas, is it only through the recognition of one's ethical/moral *responsibility* that one is justified in seeing oneself as a free subject. But this responsibility

is both *autonomous* and *universal*. Kant's insistence that "the human being . . . is subject *only to laws given by himself but still universal* and . . . he is bound only to act in conformity with his own will" (4: 432) could not be further from Levinas's equally vehement insistence that in the experience of responsibility, "no form, no capacity preexisted in me to espouse the imperative and make it my own. Not being able to treat the law as a law I myself have given myself is just in what the sense of alterity consists" (Levinas 1998a:xvii).

For Kant, "respect" is the paradigm "feeling" of responsibility for others, and this respect plays itself out negatively by not undermining the independence of others' wills from one's own (by not deceiving or enslaving or "killing" them) and positively by promoting the "happiness" of others, where happiness is both "indeterminate" (4:418) and highly individual. In these senses, Kant is entirely in agreement with Levinas about the dangers of reducing the "Other" to the "Same." Humans have a tendency, manifested in our self-conceit, to enforce our own identities by imposing them on others, thereby denying that others are unique individuals with their own goals and projects. Responsibility for the Other, for Kant as for Levinas, requires respecting others *in their uniqueness*. Moreover, for Kant, this obligation to particular others depends upon their *vulnerability*, whereby they are capable of being undermined as distinct individual wills and thus "need love and sympathy" (4:423). But for Kant this respect is due not to *anything* that is vulnerable, but only to the sorts of entities – other *people* – that are capable of governing themselves. It is by virtue of others' free capacity for choice that *they* – rather than, say, their instincts – are able to impose obligations. An adorable baby seal shares with "the neighbor" a "face" and even "the nakedness and destitution of his defenseless eyes" (Levinas 1997: 18), but the sense of responsibility one feels for that seal is, for Kant, illusory.³⁰¹ An immediate feeling of responsibility is not, even for Levinas, sufficient to establish *real* responsibility. For Levinas (as for Descartes before him), speech provides a locus to distinguish the Other (person) from simply another (thing). But Levinas recognizes that literal speech is not necessary (as the language of the "face" emphasizes) and mere speech (say, in a parrot) is not sufficient. What is both necessary and sufficient is that one recognize the other as an Other who addresses one, who calls one to responsibility. But – and here we might see Kant as providing the transcendental conditions of possibility of the "face" – this has two important implications. First, in recognizing the other, one

³⁰¹ Levinas agrees with Kant regarding humans' distinctness from animals, though Derrida does not (see Derrida 1996:71). Contemporary Kantians differ regarding humans' obligations to animals (for two important recent treatments, see Wood 1998, Korsgaard 2004).

asserts one's autonomy. It is by virtue *taking* the other to be an Other that the other can be a face. And second, the basis on which one takes the other to be an Other requires that one see the Other as sharing in common with oneself at least the ability to choose freely, to be an I that can impose responsibilities on a you. And in that sense, the "Other" is, as Levinas fears but Kant does not, reduced to the order of the Same. It is by virtue of recognizing another as being the Same (sharing one's free rational agency) that others can be "Others" (now in a Kantian rather than strictly Levinasian sense) to whom one can be responsible.

Moreover, by this Kantian standard, universal rules of justice need not be *totalizing* and thereby *violent* forms of interaction. When one seeks to enter into a community with multiple others governed by mutual respect, this *does* require subjecting individuality – one's own and those of others – to universal norms, but it does *not* require the elimination of individuality in the face of those norms. Kant explains that "pure practical reason merely *infringes upon* self-love [promotion of one's particular ends] . . . [b]ut it *strikes down* self-conceit [unconditional presumption in favor of one's particular ends]" (5:73). In a Levinasian context, we might make use of a similar distinction with respect to others. Practical reason *infringes upon* responsibilities to a particular Other in that it requires respect for every other Other, but it *strikes down* the sort of absolute responsibility implied by Levinasian and Derridian heteronomous existentialism. Put this way, it becomes clear how Levinas's philosophy helpfully draws attention to a dangerous risk of Kant's (and many other) ethics. In the recognition of common humanity and even appreciation of others' specific features, there is a risk of *reducing* personhood to shared humanity or those features. Kant, of course, insists that respect for others involves respecting and promoting individual projects and ends, but emphasizing general principles for dealing with all others can have the effect, as Levinas rightly realizes, of helping one complacency accept injustices against particular others, even against the immediate neighbor, that one ought to remedy.

At the same time, Kant serves as a valuable corrective to the enthusiastic excesses of heteronomous existentialism. For Kant, the Other – every Other – *does* have an infinite worth. One's responsibility to the Other is infinite in that one must, if necessary, sacrifice every finite interest for the sake of one's unconditional obligations. But one's obligations to any particular Other are not infinite in extent; one has specific, mostly negative responsibilities (not to deceive or torture) and general but "wide" positive responsibilities (to meet others' needs). The former are sufficiently specific that one can usually meet them merely by conducting one's behavior within certain limits. The latter, as "wide"

duties, “leave room for free choice in . . . complying with” one’s duties (6:390). Kant’s general point, against both Sartre and Levinas, is that one is not (morally) responsible for one’s whole world, nor *wholly* responsible for the Other with whom one is faced. One might think, of course, that Levinas is surely correct, that one should never succumb to the moral complacency of thinking that one has done “what one can” or “what one must” for another in need. And Kant, agrees that moral complacency is a real danger, an especially important danger given the self-deception implicit in humans’ radical evil. But Kant sees an equal danger in a “moral enthusiasm” or “moral fanaticism” that extols extraordinary moral demands but ignores the importance of ordinary duties.³⁰² After offering an interpretation of the commandment “Love God above all, and your neighbor as yourself” (5:83), Kant adds:

This consideration is intended not so much to bring clear concepts to the evangelical commandment just cited in order to reduce *religious enthusiasm* [or superstition, *Schwärmerei*]. . . but to determine accurately the moral disposition directly, in regard to our duties toward human beings . . . and to check, or where possible prevent, a *merely moral* enthusiasm which infects many people. The moral level on which a human being . . . stands is respect for the moral law By exhortation to actions as noble, sublime, and magnanimous, minds are attuned to nothing but moral enthusiasm and exaggerated self-conceit . . . they produce in this way a frivolous, high-flown, fantastic cast of mind, flattering themselves with a spontaneous goodness of heart . . . and thereby forgetting their obligation . . . If enthusiasm in the most general sense is an overstepping of the

³⁰² For an example of this sort of exaggerated moral enthusiasm consider this passage from Derrida’s *Gift of Death*:

As soon as I enter into a relation with the other, with the gaze, look, request, love, command, or call of the other, I know that I can respond only by sacrificing ethics, that is, by sacrificing whatever obliges me to also respond, in the same way, in the same instant, to all the others . . . By preferring my work, simply by giving it my time and attention, by preferring my activity as a citizen or as a professorial and professional philosopher, writing and speaking here in a public language, French in my case, I am perhaps fulfilling my duty. But I am sacrificing and betraying at every moment all my other obligations: my obligations to other others whom I know or don’t know, the billions of my fellows (without mentioning the animals...), my fellows who are dying of starvation or sickness. I betray my fidelity or my obligations to other citizens, to whose who don’t speak my language and to whom I neither speak nor respond . . .” (Derrida 1996:68-9)

bounds of human reason . . . , then moral enthusiasm is such an overstepping of the bounds that practical pure reason sets to humanity . . . If this is so, then not only novelists and sentimental educators . . . but sometimes even philosophers . . . have ushered in moral enthusiasm instead of a sober but wise moral discipline. (5: 84-6)

Kant's passion here has a practical purpose. There is a real danger that when Levinas, Derrida, and others emphasize "infinite" ethical demands to the absolutely Other in such a way that one could never, even in principle, fulfill them (or even articulate what they are!), they actually promote an even greater moral complacency than Kant's "sober but wise moral discipline" does. What seem to be intense ethical admonitions turn out to be frivolous and fantastic justifications for self-conceit.

This final Kantian critique of Levinas can also be put in terms of a more general problem with existentialist conceptions of freedom, or the absolute, or the infinite. Once freedom or the Other is distinguished from any conception that allows for discerning determinate paths of action, determinate obligations towards others, it becomes unclear what kind of normative guidance these views provide. The diversity of practical prescriptions for action amongst existentialists – from Sartre's Marxism to the Nazism of Heidegger and de Man to Levinas's emphasis on brute human suffering – only reinforces the sense that existentialism, for whatever depth of insight it might provide about the human condition, ultimately fails to answer urgent questions one asks from-within, questions about how to live and what to think. In one sense, of course, this is precisely the point. Sartre sees any attempt to "find answers" to these questions a form of bad faith, and Levinas sees the attempt to discern *determinate* ways of interacting with "others" a sort of violence against them. Rather than an ideal of *autonomy*, helping one properly govern oneself by norms, existentialists advance an ideal of *authenticity*, accepting the normlessness of the from-within and refusing to subject oneself or others to the violence of being forced into determinate categories of thought and action. So it is, at last, to a Kantian account of authenticity that we must now turn.

Kantian Authenticity

Both Kant's privileging of objective knowledge and his emphasis on the moral law as the law of freedom seem, from an existentialist standpoint, to be forms of inauthenticity. Nietzsche puts the point with

suitable vitriol, criticizing the herd mentality involved in preferring “truths” that can be *shared* over the “secret gardens” (Nietzsche 1967:17) that one can cultivate for oneself. But Heidegger’s analysis of “the they,” Sartre’s account of bad faith, and the Derridean-Levinasian critique of “Reason” all point to the same general criticism. Kant seems to be the inauthentic philosopher extraordinaire, offering transcendental conditions of the possibility of thinking-just-like-everyone-else and acting-just-like-everyone-else. In a move that would be comic were it not so tragic, Kant even seeks transcendental conditions of possibility of universal judgments about beautiful objects, finding universality even in humans’ only truly “*free* satisfaction” (5:210). Against Kant’s attempts to justify necessary and universal standards of thought, feeling, and choice, the existentialist ideal is a call to authentically think, feel, choose, and *be* for oneself.

To some extent, Kant would vehemently endorse the (early Heideggerian-Sartrean) existentialist ideal of authenticity. Kant’s “What is Enlightenment?” famously begins with the claim that “Enlightenment is the human being’s emergence from his self-incurred . . . inability to make use of his own understanding without direction from another” and Kant mocks those who “have a book that understands for me, a spiritual advisor who has a conscience for me . . . and so forth” (8:35). In his account of character, so fundamental to both his moral and empirical anthropology, Kant insists that “the *imitator* . . . is without character; for character consists precisely in originality in the way of thinking. He who has character derives his conduct from a source that he has opened by himself” (7:293). And the conceptions of autonomy that play such a central role in Kant’s transcendental anthropology are based on the fundamental principle that “the human being . . . is subject *only to laws given by himself* and . . . is bound only to act in conformity with his own will” (4: 432). In these contexts, Kant accepts and even emphasizes the fundamental existential commitment to one’s *ownmost* possibility, to an authenticity that refuses to see its own choices as the choices of simply a generic “human being.” Kant’s conception of “autonomy” is precisely designed to capture this important commitment to what existentialists call authenticity.

But this emphasis is combined, in Kant, with a commitment to *universality*. “What is Enlightenment?”, even while defending “one’s *own* understanding” (8:35), insists that only a “*public* use of one’s reason” brings enlightenment (8:35, 37). In discussing character, Kant not only contrasts true character with “the imitator” but also adds, “However, the rational human being must not be an *eccentric*; indeed, he never will be, since he relies on principles that are valid for everyone” (7:293). And his

insistence in *Groundwork* that one be bound only to laws given to oneself adds that these laws must be “still universal” (4: 432). Kant’s emphasis on autonomy and thinking for oneself is *not* an emphasis on *uniqueness*. One’s ownmost possibility, for Kant, is not a possibility that sets one apart from others, but a possibility that one shares with others. And Kant vehemently protests against “egoisms” of various kinds, wherein one takes one’s own beliefs, feelings, or desires to be sufficient without testing them against “other men’s insights,” which Kant calls the “touchstone of the understanding” (24:178, see too 5:294, 7:128-9).

Moreover, Kant argues that human autonomy is not merely negative, not something that arises from Nothing or from an encounter with an inscrutable Other. Autonomy has a positive structure. When one considers what is necessary in order to truly pursue one’s ownmost possibilities, one finds that certain determinate and universal laws must govern one’s thoughts and volitions. Against existencialisms of all sorts, Kant argues that there can be and in fact *is* a single categorical imperative by which all *authentic* agents govern themselves. The fact that there is such an imperative, and that it can be known “a priori,” might seem to imply that this moral law is distinct from and supreme over one’s freedom. But for Kant, subjection to the moral law is precisely the condition of possibility of having freedom at all, and thus a condition of possibility of *any* sort of authentic engagement with the world. The fact that this law is the same *for everyone* seems to make Kantian morality a paradigmatic “herd morality,” a morality of “the they” rather than a truly authentic expression of one’s *ownmost* possibilities. Here again, though, it is important to recall that existentialist authenticity is not as much about what one does as about how one does it. And for Kant, as for existencialists, one ought not obey the moral law simply because others are doing so, nor simply because Kant says one should, nor even because it is a law of God. One ought obey the moral law because it is the law of one’s own will. Thus even if the moral law is identical to the law of “the they,” the Kantian moral agent will obey it *as* a law of her own will. She will obey this law authentically. Moreover, Kant is well aware that particular judgments of particular societies – of “the they” – can be at odds with the moral law. The fact that Kant’s ideal moral (and even epistemic) standards are normatively universal (in that they ought to hold for all people) does not mean that they are universal “norms” in the sense of actually being embraced by “the they.” By providing (universal) conditions of justification from-within for individual agents, Kantian autonomy can even supply emboldening resources to stand up authentically against the “norms” of one’s society.

In their appeals to authenticity and warnings about succumbing to “bad faith” or “herd mentality,” existentialists rightly highlight concerns that can be missed by a Kantian focus on autonomy. Despite Kant’s insistence on autonomy being a form of *self*-governance, it can often become merely a stale formula in the context of which one lives one’s life. One can treat the moral law as a constraint given from on high, or as a social norm to which one must conform one’s life. Kantians should emphasize the importance of always obeying the moral law *as one’s own law*, a law that, although necessary, is nonetheless imposed upon oneself freely. But Kant, in return, can enrich existentialism by highlighting that taking seriously the existential imperative to authenticity cannot be merely a matter of asserting uniqueness or constantly reinventing oneself. To say, as Sartre’s student said about joining the Resistance, “In the end, feeling is what counts. I ought to choose whichever pushes me in one direction” (Sartre 1993: 25), is a far cry from authenticity. *True* authenticity comes, as Kant insists, from *autonomy*, from governing oneself in accordance with laws that one can see as truly coming from one’s *self*, laws that will, by virtue of being *laws* and thereby susceptible to justification, be universal.

Within the broad universal laws by which we govern ourselves, Kant admits the importance of individuality. The moral law is not as *radically* adverbial as the existentialist ideal of authenticity; it implies some content, such as that we must avoid false promises and promote the perfection and happiness of self and others. But it is still a *formal* law within which there is considerable range of possibilities, many (incompatible) ways of being moral, and thus lots of room for particular expressions of one’s “ownmost” possibilities. There is room, even within Kant’s own system, for authentic actions that are universalizable in principle (and hence moral) but still expressions of one’s uniqueness. Sartre’s own example illustrates this point particularly well. Sartre rightly notes, in the case of the student deciding whether or not to join the French Resistance, that Kant’s ethics does not help him decide whether to join or to stay with his mother. But Sartre is wrong about the *reason* why Kant’s ethics fails. As Sartre sees it,

if I stay with my mother, I'll treat her as an end and not as a means; but by virtue of this very fact, I'm running the risk of treating the people around me who are fighting, as means; and, conversely, if I go to join those who are fighting, I'll be treating them as an end, and, by doing that, I run the risk of treating my mother as a means. (Sartre 1993: 25-6)

Sartre suggests that one violates one's moral responsibilities whatever one does, but in fact, the problem is quite the opposite. The relevant duties (to his mother, to the Resistance) are imperfect ones. One must be responsive to imperfect duties, but one is not strictly required to fulfill any particular ones. Thus the student can be autonomous whichever of these options he chooses. This does not, however, mean that "authenticity" is *all* that matters. The moral law restricts the *range* of choices. He cannot deceive his mother for the sake of her feelings, telling her that he is going to a desk job when he will really be going to the front. He cannot sign a pledge to join the Resistance and then back out when the going gets tough. He cannot join the Nazis and fight against the Resistance. But within the limits of the norms that he ought to impose on himself, there is room for him to decide his particular destiny.

Here again, Kant has something important to learn from existentialism. Although Kant's empirical account of human beings includes a wide range of different propensities and inclinations, his from-within treatment of volition tends to see all choice as essentially just a matter of how one prioritizes happiness/self-love vis-à-vis duty. What existentialists show is that particulars are significant not merely as ways to pursue happiness or do one's duty. When it comes to choosing *amongst* morally acceptable ends, Kant's emphasis on happiness (both oneself and others') and perfection (especially one's own) does not sufficiently recognize the importance of manifesting freedom even with respect to not-strictly-moral choices. Choices of particular paths in life come to define who we take ourselves to be, and even if only morality matters "without qualification," particular concerns can be much more closely tied to our self-conceptions than their merely moral dimensions suggest. Existentialism adds that much of what makes human life worth living cannot be reduced to formulae, that one must make choices about what to think, feel, and do that cannot be dictated by rules. Or, rather, that insofar as one does choose what to think, feel, and do in accordance with rules, one is necessarily blind to something – whether, as for Sartre, something *in oneself*; or, as for Levinas, something *in others* – that transcends those rules. This attention to what cannot be reduced to rules, categories of the understanding, or publically-accessible language and concepts reflects a genuine contribution of broadly existentialist approaches to the self, one that Kantians would do well to take seriously.

That said, Kant also has important things to teach existentialists. Even if universal rules are limiting, one must still think and act in the world. And knowing merely that one's rules and concepts underdetermine one's thoughts and choices does not actually help one make choices. By providing a priori rules for genuinely free thought and

action, Kant's philosophy meets problems real thinkers and choosers face. For one thing, Kant's transcendental analysis of volition helpfully reveals a fundamental difference in urgency between two sorts of choices. As important as it might seem – and really be – to choose the right partner(s) in love, the right career(s), and the right balance of work, deep relationships, and raw pleasures, these choices are less important than choices of whether to do what is right or to forego moral requirements for the sake of other ends. Moreover, Kant rightly points out that even within the structure of morality, there is important room for choice. One has imperfect duties to perfect talents and promote others' welfare, but this general structure of duty does not tell a person whether to become a concert pianist, a subsistence farmer, a work-a-day paralegal with a vibrant social life, a stay-at-home-parent, or a professional philosopher. Which career one chooses will have immense impact on one's sense of oneself and the quality of one's life. But Kant insists that *all* lives constructed in the context of an overall commitment to duty are good without limitation; any such life constitutes the supreme good for human being, irrespective of its specific form. And while this realization does not eliminate the angst that can be faced by those making difficult life decisions, it puts those decisions in a larger context that can help one avoid paralyzing despair over one's choice and/or haughty hubris about the importance of one's chosen path.

In the end, Kant can and should embrace much of the existential "ideal" of authenticity while offering a valuable antidote to the cult of authenticity prevalent today. For Kant, something like authenticity is extremely important. Autonomy, like authenticity, involves governing oneself, refusing to let one's decisions be dictated by the "they," making choices with awareness of one's own freedom. And for all of the content (albeit formal) that Kant gives his conception of autonomy, Kant – like existentialists – insists that merely following the categorical imperative is insufficient to be autonomous; one must do what is right *autonomously*, that is, one must will universally and respect the humanity of others *because one legislates this law to oneself*. For Kant, one who does what is morally required because it is required by God, or one's society, or one's ultimate happiness, or, frankly, for any reason other than as a form of *self-command*, does not act autonomously and thus does not even really do what is morally required. In that sense, Kant's ethics of autonomy is an ethics of authenticity. And Kant can and should enrich his account with a greater emphasis on choosing authentically, in a way that is genuinely responsive to who one takes oneself to be, in every area of life, even when morality is not specifically at issue. But for Kant, authenticity is limited in that one can be *truly* authentic only insofar as one acts

morally, and choosing morally well is the only sort of authenticity that matters *unqualifiedly*. But these limitations, far from weakening Kant's overall account, make Kantian existentialism more realistic, more just, and more effectively action-guiding than its more radical counterparts.

Chapter 11: Normativity

The last three chapters looked at several approaches to thinking about human beings that foundered on problems related to normativity. As a result of work in the biology and psychology, we understand better than ever how the human mind – as an empirical object – works. Through study of human history and, more generally, of human diversity, it is increasingly clear that there are wide variations in the ways human beings think and govern action. And philosophical developments over the past 100 years, especially existentialism, have challenged traditional dualisms and driven home the difficulty of determining any fixed human “nature.”

Even with all of these developments, however, normative issues have not gone away. Insofar as human sciences (whether biology or history or anthropology) purport to provide anything like knowledge, they appeal to epistemic standards. Moreover, as forms of life, the practice of these sciences involves appeal to volitional standards. Even those not directly involved in these scientific and philosophical development need – at least now that you have read this far in this book! – to decide what to make of them. And in any case, as reflective thinkers, feelers, and actors, we find ourselves – again, especially if you’ve read this far – needing to decide how to think, feel, and choose. As existentialists rightly show, even refusal to see oneself as needing to make such decisions is an exercise of freedom, albeit in bad faith.

Unfortunately, none of these developments provide, in themselves, the tools for deciding how we ought to think, feel, and act in the light of them. Even if evolutionary biology shows how and why humans evolved to care about one another, this does not in itself help us determine whether such caring is a trait to be fostered or resisted. And even if historical analyses show how modern physics developed through the overturning of a Newtonian paradigm, how modern conceptions of penal justice developed, or even how human beings as subjects emerged, historical analysis alone cannot tell us whether to believe the claims of modern physics, embrace current conceptions of penal justice, affirm modern subjectivity, or reject one (or all) of these developments. Comparison of different ways of thinking and acting across lines of gender, race, or culture, likewise, can reveal the contingency of these ways of thinking, but it cannot tell us whether this contingency should be embraced as a delightful pluralism, rejected as tribal ethnocentrism, or treated in some wholly different way. And even existentialism itself, with its emphasis on freedom, so emphasizes freedom and authenticity, and calls into such great suspicion any purportedly universal normative standards, that it fails (and intentionally so) to provide real guidance for

making the important decisions the omnipresence of which it so aptly reveals.

As we saw in chapter one, Kant's transcendental anthropology provides a normative framework for answering questions about what to think, feel, and do. Kant's purely formal normativity leaves room for individual thoughts and choices to be influenced by historically-evolving cultures and biological capacities/tendencies while still being incorporated into normative frameworks that are at once authentically one's own and genuinely universal for human beings. But Kant has hardly been the last word in thinking about how human beings ought to think, feel, and choose. In chapters six and seven, we saw a wealth of alternatives emerge in century after Kant. The past century has seen an equally rich range of normative theorizing informed by increasingly rich conceptions of human beings. In fact, as natural and human sciences became more rigorous and informative, they split from "philosophy" to form their own disciplines, leaving philosophers the normative issues that sciences are incapable of addressing themselves. As a result, contemporary philosophy – insofar as it is not merely existential reverie or handmaiden to science – focuses on human beings from a normative perspective.

To summarize all of contemporary philosophy in a single chapter is impossible, so this focuses on just five recent and important approaches to normativity. With the exception of the first, these approaches all aim, in one way or another, to take seriously the problems posed by naturalism, existentialism, and human diversity. In that sense, they all speak, in a more direct way than Kant, to our contemporary situation. I first discuss intuitionism, a cluster of philosophical approaches that dominated much 20th century philosophy. I then turn to Rorty's pragmatism and MacIntyre's emphasis on narrative and tradition, two provocative recent approaches to normativity that seriously address naturalism, historicism, and existentialism. The last two approaches – Habermas' communicative ethics and Korsgaard's neoKantianism – are arguably the most influential recent appropriations of Kant in contemporary moral philosophy. Both incorporate substantial elements of naturalism, historicism, and existentialism while remaining recognizably and self-consciously Kantian. They thus provide excellent though controversial examples of how Kant might be updated for the 21st century.

I. Phenomenology and intuitionism, then and now

We start our grand tour of contemporary philosophy with two early 20th century thinkers whose general approaches, though not always explicitly invoked, still dominate much contemporary philosophical theorizing: Edmund Husserl (1859-1938) and G.E.Moore (1873-1958).³⁰³ Both thinkers, in different ways, advocated the use of “intuitions” to discern philosophical truth, and both left extensive legacies on contemporary philosophy. One result is that much contemporary philosophical work is unabashedly intuition-driven.

Husserl’s phenomenology represented an early attempt to reinvigorate philosophy in the face of a rising “psychologism” by focusing philosophical research on the search for “essences.” Husserl sees “the question of philosophy’s relation to the natural and humanistic [i.e. historical] sciences” as the defining problem for philosophy and aims to establish philosophy “as a rigorous science” without falling into scientific naturalism (Husserl 1965:72). Against the neoKantianism that dominated German philosophy in the early 20th century and preached “back to Kant,” Husserl proposes that philosophers go “back to the things themselves” (Husserl 1965:96-7), where the “things” at which Husserl aimed are not individual objects but *essences*: “Pure phenomenology as science . . . can only be essence investigation” (Husserl 1965:116, cf. Husserl 1983:8, 65-6). While natural and human sciences might study relations between particular things in the world, they can never make the necessary and universal claims about things that constitute claims about something’s *essence*. And while Kant focuses on merely *formal* structures of experience, Husserl insists on a “material” a priori that goes far beyond Kant’s own transcendental philosophy: “nature with all its thing-like contents certainly also has its a priori” (Husserl 2001:43) such that, for instance, we can know a priori that “colors” cannot literally be heard as such; such a claim is based on an insight into the fact that it is essential to being a “color” that it is specifically *visual*.

One gets access to essences by means of what Husserl calls “eidetic seeing” or “essential intuition” (Husserl 1965:110, Husserl 1983:8-9) and such intuitions, properly understood, provide a “foundation free of doubt” (Husserl 1965: 76) for philosophical

³⁰³ Husserl is widely seen as the father (or grandfather) of contemporary “Continental” philosophy, an approach dominant in France and Germany that, today, largely emphasizes the sort of existentialist philosophizing of figures like Derrida and Levinas. Moore, along with Frege and Russell, is one of the fathers (or grandfathers) of what is generally called contemporary “analytic” philosophy, dominant in the English-speaking world.

progress.³⁰⁴ While “intuition” is infallible as a foundation, its reliability depends upon remaining “*within the limits in which it is presented there*” (Husserl 1983:44). With respect to this limiting, Husserl insists that one focus on “phenomena” as presented to consciousness, which requires “a general reversal of that ‘natural attitude’ in which everyday life as a whole as well as the positive sciences operate. In [this natural attitude] the world is for us the self-evidently existing universe of realities which are continuously before us in unquestioned givenness” (Husserl 1981:27b). Husserl describes this “reversal” of the natural attitude as an “*epoche*,” a “bracketing” of ordinary assumptions about the world in order to focus specifically on what is immediately present in intuition. By bracketing assumptions about individual objects, one can effect an “all-embracing transition from the factual to the essential form, the *eidōs*” (Husserl 1981:25b). One key way of doing this is through “imaginative variation”:

If the phenomenological actual fact . . . serves only as an example and as the foundation for a free but intuitive variation of the factual mind and communities of minds *into* the a priori possible (thinkable) ones; and if now the theoretical eye directs itself to the necessarily enduring invariant in the variation; then there will arise with this systematic way of proceeding a realm of its own, of the “a priori”. There emerges therewith the eidetically necessary typical form, the *eidōs*; this *eidōs* must manifest itself throughout all potential forms of mental being in particular cases. (Husserl 1981:25b)

Rather than asking about how and when people *actually* become aware of colors (a sort of factual investigation of color), one can take the experience of a particular color as an occasion for a “free but intuitive variation” (Husserl 1981:25b, see too Husserl 1983:157f.), wherein one imaginatively runs through an indefinite number of variations on “color,” intuiting what is essential to color as such, and getting “knowledge which will have validity far beyond the psychologist’s own particular psychic experience” (Husserl 1981:23b).³⁰⁵

At this level of generality, Husserl’s phenomenology can be applied to anything: tones, colors, material objects, and so on. But the primary focus of Husserlian phenomenology is consciousness itself, that is, to

³⁰⁴ Though see Husserl 1983:212, where Husserl admits the possibility of “errors” in phenomenology, but then proposes “weeding out errors by measuring them against intuition.”

³⁰⁵ By bracketing the natural attitude, one considers possibilities not in terms of possible realization *in the objective world* but in terms of the essence of color as such. Even if, scientifically speaking, color is impossible without “energy,” one can eidetically distinguish between colors and various interpretations of energy (as possible work, possible heat, etc) by imaginatively considering a possible experience of color independent of any energy-effects.

discern the essence of consciousness in its various forms (perception, recollection, valuing, etc.): “the universal task . . . : to investigate systematically the elementary intentionalities . . . and from this advance towards a descriptive knowledge of the totality of mental processes, toward a comprehensive type of a life of the psyche” (Husserl 1981:23b, see too Husserl 1983:67ff.). From this methodology and focus, Husserl developed an elaborate phenomenology of consciousness. His most important eidetic insight was the recognition that consciousness is essentially “intentional,” in that it “intends,” or is directed towards, particular objects. That is, consciousness in all its forms is consciousness of ___ (perception of, valuing of, willing of, etc). From this, Husserl develops an elaborate account of what he calls “noesis” and “noema,” referring (respectively) to the “intentional mental process” by which an object is intended in a particular way (e.g., one perceiving of a cup of coffee) and the objects as objects of consciousness (the coffee as perceived). (Again, because the natural attitude is bracketed, even “imaginary” objects are noema, in the sense that one’s imagining of a unicorn has an object just as much as one’s perception of a cup of coffee.)³⁰⁶ Both Husserl’s general methodology and his specific claims have been appropriated and critiqued by an ongoing phenomenological tradition that aims to discern essential structures of consciousness through an examination of them from-within that brackets, if not the natural attitude as such, at least the scientific-naturalistic assumptions that can interfere with paying close attention to consciousness. Existentialism itself, as we saw in the previous chapter, is the most prominent representative of this phenomenological tradition.

G.E.Moore operated within a very different philosophical climate than Husserl but articulated views that, in important respects, are quite similar. Moore’s most famous philosophical position is his attempted “Proof of the Existence of the External World.” Moore claims to give a “perfectly rigorous” proof of the external world by showing that two things (specifically two hands) exist in the external world:

How? By holding up my two hands, and saying, as I make a certain gesture with the right hand, ‘Here is one hand’, and adding, as I make a certain gesture with the left, ‘and here is another.’ (Moore 1959:146)

³⁰⁶ For discussion of noesis and noema, see Husserl 1983:205ff. (through the end of volume 1).

Moore raises the obvious objection, that he does not really “know” that his hands exist, and responds:

I *knew* that there was one hand in the place indicated by combining a certain gesture with my first utterance of ‘here’ and that there was another in the different place indicated by combining a certain gesture with my second utterance of ‘here’. How absurd it would be to suggest that I did not know it, but only believed it, and that perhaps it was not the case! You might as well suggest that I do not know that I am now standing up and talking — that perhaps after all I’m not, and that it’s not quite certain that I am! (Moore 1959:146)

For Moore, the reason the proof *seems* inadequate is that there is no “proof” of his premises, “Here is one hand” and “Here is the another.” But, Moore argues, it is not necessary to *prove* these premises, as long as one *knows* them, and “I can know things, which I cannot prove; and among things which I certainly did know, even if (as I think) I could not prove them, were the premises of my two proofs” (Moore 1959:148). Strictly speaking, Moore does not describe these known-but-not-proven premises as “intuitions,” and he is here better understood as a commonsense philosopher than an “intuitionist” strictly speaking.³⁰⁷ But this proof captures the spirit of building one’s philosophy on the basis of obviously true claims, and in that sense, contributes to a broader sort of intuitionism in recent philosophical approaches to normativity.

Moore’s worked-out philosophical intuitionism comes in his ethics. Within ethics, Moore argues against attempts to define the good or identify it with any other property (such as happiness or universalizability). Moore accuses all such attempts of a “naturalistic fallacy,” which, in its most common form, involves the identification of “the good” with some natural property such as pleasure. Moore argues that for any such definition, it remains an “open question” whether “pleasure (or whatever it may be [that one identifies with the good]) [is] after all good?” And one “can easily satisfy himself that [in asking this question] he is not merely wondering whether pleasure is pleasant” (Moore 1913:16). In the case of Kant, it seems coherent to ask, “Is acting only on universalizable maxims good?” As anyone who has taught Kant’s ethics knows, this question is notoriously unlike the question, “are universalizable maxims universalizable?” Countless people answer the former question with a resounding “No” when considering such tricky

³⁰⁷ Moore is also very far from *Husserlian* intuitionism because he assumes the natural attitude and uses common sense to make substantive existential claims (not mere claims about essences and certainly not claims limited to a phenomenology of consciousness).

cases as the Nazi at the door. For Moore, this implies that “‘good’ denotes a simple and undefinable quality,” much like basic perceptual qualities such as “yellow” (Moore 1913:10).

Because goodness is simple and undefinable, Moore turns to “intuition”³⁰⁸ to discover which states of affairs have “goodness”: “[we] must consider what things are such that, if they existed by themselves, in absolute isolation, we should yet judge their existence to be good” (Moore 1913:187). Like Husserl, Moore emphasizes imaginative variation to focus intuition:

[I]f we place before ourselves the question: Is consciousness of pleasure the sole good? The answer must be: No . . . [W]e isolate consciousness of pleasure [and] ask: Suppose we were conscious of pleasure only, and of nothing else, not even that we *were* conscious, would that state of things . . . be very desirable? No one, I think, can suppose it so. (Moore 1913:95)

For Moore, the question “What ought I to do?” reduces to “What kind of things ought to exist for their own sakes?” and “What things are related as causes to that which is good in itself?” (Moore 1913:viii, 146). The first is answered via intuitions that certain states of affairs have the simple and undefinable property of goodness; the second via ordinary, scientific causal reasoning (based, among other things, on our commonsense knowledge of objects in the external world). Thus intuitions form the core of Moore’s account of volitional normativity.

Neither Husserlian phenomenology nor Moorean intuitionism are particularly popular today. Both Moore and Husserl suffered philosophical attacks from former students – Heidegger in the case of Husserl, Wittgenstein in the case of Moore – attacks that affect how phenomenology and intuitionism have been appropriated today. In both cases, these attacks targeted (among other things) how Moore and Husserl sought *unrevisable* conclusions based on *isolating* particular intuitions and both Heidegger and Wittgenstein rejected the broadly *foundationalist* philosophical program shared by Husserl and Moore, in which indubitably self-evident intuitions provided a foundation for a scientifically rigorous philosophical system. Most contemporary philosophers follow Heidegger and Wittgenstein in giving up any attempt

³⁰⁸ Note that while Moore argues that “intuitionism is mistaken” (Moore 1913:148), he has in mind only intuitionism about particular duties; his own view is a form of intuitionism about goods, from which one gets duties by using knowledge of natural connections of causes with those goods as effects.

to derive an indubitable philosophical system on the basis of intuitions gleaned from isolating particular phenomena.

However, important features of these thinkers' approaches are widely shared within epistemology, moral philosophy, and even metaphysics. The phenomenological appeal to what can be made evident through careful attention to engagement with the world is evident, for example, in the existentialist phenomenologies of Heidegger, Sartre, and Levinas. But the importance of intuition is particularly prominent in contemporary normative theorizing in epistemology and moral theory, especially among philosophers who aim to avoid or supplement scientific naturalism and who have neither concern for historicism nor sympathy with existentialism.

One place where the appeal to intuition has been particularly prominent is within the field of "analytic epistemology," a philosophical school focused on "a priori theorizing about the nature, conditions, and extent of human knowledge, rationality, and justification" (Sosa 2009:103).³⁰⁹ The most discussed problem within 20th century analytic epistemology is based on our "intuitions" about what should count as "knowledge." A dominant account of knowledge posited that knowledge could be identified with "justified, true belief" (e.g. Chisholm 1957:16). But in 1963, in his now famous "Is Justified True Belief Knowledge?", Edmund Gettier proposed two cases (many more would follow) that "show" that knowledge cannot merely be justified true belief. One case involves supposing,

Smith has strong evidence . . . [that] Jones owns a Ford . . . [and infers that] (h) 'Either Jones owns a Ford, or Brown is in Barcelona.' Smith is . . . completely justified in believing [this] proposition [since it follows from the claim that Jones owns a Ford]. But imagine now that . . . Jones does *not* own a Ford, . . . [and], by the sheerest coincidence, and entirely unknown to Smith, . . . [Barcelona] happens really to be the place where Brown is. If these two conditions hold, then Smith does *not* know that (h) is true, even though (i) (h) is true, (ii) Smith does believe that (h) is true, and (iii) Smith is justified in believing that (h) is true. (Gettier 1963:123)

³⁰⁹ Strictly speaking, this is only one of the dominant approaches within contemporary analytic epistemology, the other being a "more naturalist epistemology that studies contingent ways in which we humans can and do satisfy conditions revealed by reflection as necessary for human knowledge of one or another variety" (Sosa 2009: 103). Since I discussed naturalist epistemology briefly in chapter eight, I focus my discussion here on more classical analytic epistemology.

Like Husserlian phenomenology, this sort of epistemology involves imaginatively varying conditions to better intuit the essence of something, in this case “knowledge.” The conviction that this is not a case of real knowledge is based entirely on the intuitions of the epistemologists studying it. If one thinks that Smith *does* know (h), there seems little reason to challenge the conception of knowledge as justified true belief, and thus little sense in much contemporary epistemology, which aims to respond to Gettier-style counter-examples to various definitions of knowledge.

The general style of argument for analytic epistemologists is, as in the case of Husserl and Moore, intuition-driven. Just as Husserl sought to discern “essences” by means of imaginative variation, contemporary analytic epistemologists continually form and refine accounts of belief, knowledge, justification, and so on based on intuitions formed about countless examples and counter-examples. The goal is to get to universal and necessary (a priori) claims about *the* nature of knowledge, belief, and so on; and thus (in ways that are increasingly suspect³¹⁰) to get an answer to the question of what one ought to believe.

Similarly, within contemporary moral philosophy, much work is done by moral intuitions.³¹¹ “Moral intuitionism” of the Moorean sort is largely rejected by contemporary moral philosophers, who, if they are not Kantians, prefer various kinds of “utilitarianism” or “virtue ethics.” But within all three major approaches to ethics, moral intuitions often play major roles. Kantians make use of intuitions both in establishing key starting points (such as the claim that moral claims must be universal) and, often, in discerning what sorts of maxims can actually be willed universally. Utilitarians often appeal to intuition as the basis of their principle of utility. Mill’s claim that happiness is the only good is based on a series of intuitions that in everything else we want, we ultimately want happiness. Virtue ethicists, likewise, end up appealing to intuitions either about particular “thick” ethical notions (courage, temperance, etc) or about what constitutes human flourishing. Finally, even insofar as epistemology or moral philosophy shift from using intuitions for *foundational* roles, intuitions are still *accommodated* by through *coherence* or *reflective equilibrium*.³¹² Almost all contemporary ethical theory appeals, at least implicitly, to a process of “reflective equilibrium”

³¹⁰ See e.g. Sosa 2009, which admits that “knowledge” might not even be the highest purely epistemic value.

³¹¹ See Frazier 1998, Jones 2005, Sinnott-Armstrong 1992, Sturgeon 2002.

³¹² For the classic discussion of reflective equilibrium, see Rawls 1971:20. For recent discussions of coherentism, Brink 1989, Sayre-McCord 1996, and Timmons 1999.

whereby moral theories are susceptible to criticism insofar as they diverge too significantly from common sense moral beliefs. Thus, famously, divergence between moral intuitions about lying to a murderer at the door and Kant's rigorist adherence to truth-telling is taken as a serious challenge to Kant's moral theory.

Recently, intuitionism in philosophy has been heavily criticized from at least two perspectives. First, intuitionism has been criticized from those (whom we might see as broadly existentialist) who simply ask why one should take the fact that something *seems* true to mean that it *is* true, those, that is, who recognize that in general one is capable of standing back from one's intuitions and asking about them, "But is this *really* true?" Christine Korsgaard, for example, points out that intuitionists simply fail to answer normative problems precisely when they are most acute: "if someone falls into doubt about whether obligations really exist, it doesn't help to say, 'ah, but indeed they do. They are *real* things.' Just now he doesn't see it, and herein lies his problem" (Korsgaard 1996: 38). Human beings can – and often must – ask themselves whether the values that seem to have mattered to them really *do* matter, and the intuitionist who says, "Think carefully and it'll be obvious" fails to address the human situation of *not* finding the value of values obvious. Second, naturalists, historicists, and those who emphasize human diversity point out that intuitions are largely rooted in culture. As ways of figuring out the nature of "truth" as such or of discerning what we "ought" to believe in some absolute sense, it seems unreasonable to favor our (contingent) intuitions over those of people with other backgrounds. Even for more coherentist approaches, there is a standard "garbage-in-garbage-out" objection; if one's initial set of intuitions is misguided or historico-culturally contingent, one will reach what is only a misguided or historically local equilibrium, one that gives little grounds for believing that one has discovered normativity in any supra-historical sense. Such critics argue that what intuitionist epistemologists and moral philosophers are really doing is offering a highly local "anthropology" that merely specifies the carefully taught (and selected for) prejudices of a small tribe of mostly-wealthy, mostly-white, mostly-male philosophers in English-speaking universities in the 20th and early 21st centuries. As Stephen Stich has quite dramatically put the point: "[W]e think that the best reaction to the High-S[ocio]E[conomic]S[tatus], Western philosophy professor who tries to draw normative conclusions from the facts about 'our' intuitions is to ask: What do you mean 'we'?" (Weinberg et. al. 2008: 40, see too Appiah 2009, Knobe 2008).

Kant's philosophical approach can, I think, help with both sets of concerns. With respect to naturalist and historicist critiques, three points should be made, related to similar points made in defense of Kant in chapters eight and nine. First, naturalist and historicist critics of intuitionism simply mistake the important difference between the transcendental and empirical perspectives, or between linguistic analysis and intuitions about the realities to which one's words apply, e.g., "the being square of any figure is a different condition from its being . . . 'square' in my idiolect" and "the question . . . whether [the term] 'knowledge' applies to the protagonist in an example" is not the same as "the question [of] whether the protagonist . . . *know[s]*" (Sosa 2009: 105, 104).³¹³ Second, while there may be considerable historical and cultural divergence with respect to certain forms of human life, there is considerably less divergence with respect to others, such as the experience of an objective world and the constraint of personal choice by moral norms. Third, as we noted in chapters eight and nine, the very naturalist and historicist claims that seem to call our intuitions into question are based, at least to some degree, on those very intuitions. Without some reason to view as reliable the scientific methodology behind comparative studies of human intuitions, there is no reason to consider any results of such studies as calling into question one's intuitions. To *some* degree at least, one must take one's own perspective, one's own sense of what counts as justification or knowledge, for granted if one is to take historicist or naturalist objections seriously. Depending upon the results of such studies, one may have to give up confidence in the legitimacy of *some* intuitions, but one cannot give up confidence in them *all* without giving up the studies themselves.

With respect to more existentialist concerns, Kant might rightly point out that our "standing back" from intuitions is always a *normative* standing back, that is, a way of asking whether one *should* remain committed to what one finds oneself committed to. And normatively standing back depends upon standards for normative evaluation, so that one cannot stand back from intuitions without having or at least seeking some basis for evaluating them. This does not mean that one can evaluate intuitions only based upon other intuitions one already has, nor that one must adopt a method of reflective equilibrium. In that sense, *intuitionism* may turn out to be the wrong way to go about developing a normative theory. But it does imply that *some* sort of normativity is

³¹³ For more detail and for various objections to Stich and similar critics of intuitionism, see Murphy and Bishop 2009.

necessary in order even to stand back from the intuitions one finds oneself to have; in the absence of other candidates, this requires evaluating some intuitions on the basis of others.

As these responses might suggest, however, Kant's philosophy also differs in important ways from contemporary appeals to intuition. Throughout Kant's transcendental anthropology, he appeals to very general intuitions such the existence of obligations, the fact that we experience an objective world, the impossibility of imagining this world non-spatially, and so on. Kant's own defense of the categorical imperative, for instance, does not depend upon intuitive awareness of that particular moral law, nor even a process of reflective endorsement whereby the law is confirmed by its conformity with our moral intuitions. *Kant's* argument is – or at least aims to be – simply that if we have *any* moral intuitions at all, that is, any sense that we are obligated to do anything, then we are committed to a sort of moral responsibility the condition of possibility of which is that we are bound by this particular moral law. Thus Kant avoids brute appeals to particular intuitions while still providing some place for intuitions of the most general sort.

II. “Cheerful Ethnocentrism”

Ordinary people as well as philosophers continue to use intuitions to decide what to think and do and what theories of knowledge and action to adopt, but the historical and cultural variability of intuitions and the seeming impossibility of any truly universal, foundationalist philosophical system has led to alternative approaches to normativity. One such approach accepts that one's intuitions are simply one possible set of intuitions among others but embraces them nonetheless. One of the most important and interesting contemporary approaches along these lines is a form of neo-pragmatism espoused in its most influential form by Richard Rorty.

We can understand Rorty's fundamental claim in terms of the Kuhnian, historicist account of science in chapter eight. Rorty generalizes Kuhn's claim that “normal science” is characterized by adhering to rules and norms intrinsic to a particular paradigm and Kuhn's rejection of fixed and determinate standards for evaluating the revolutionary science involved in fundamentally changing paradigms. For Rorty, every “language game” is governed by norms for the use of words within that game, and in that sense, every ordinary use of language is akin to the practice of “normal science.” Rather than a specifically

“epistemic” normativity that would ask “What should I believe?”, Rorty focuses on a more linguistic normativity that asks, “What should I *say*?” (which will include such things as saying to oneself, “I believe such and such”). And the answer to that question is typically local. Language-games include their own norms, and one should say only what is consistent with the internal norms of one’s game. For Rorty, the primary sense of normativity is a matter of “criterion-governed sentences within language games” (Rorty 1988: 5). Within certain religious language-games, it makes sense to say that “Phoebus rounded the earth in his flaming chariot” or that “[God] the golden-tressed sun, makes his daily course to run.” Within the ordinary language of daily life and weather channels, it makes sense to say, “Tomorrow, the sun will set at 7:13 PM.” And in the language-game of introductory astronomy, it makes sense to say “The earth lies at a distance of 150 million kilometers from the sun, orbiting the sun at a speed of 100,000 kilometers per second, and spinning on its axis with one full rotation every 23.93 hours.” For all these cases, criteria of correctness are *internal* to the particular language game,

One can, of course, ask whether one ought to adopt a particular language game. If one chooses to use a particular term – say, “spinster,” “heretic,” “homophobe,” “phlogiston,” “quark,” or “essence” – then the use of such terms is governed by norms intrinsic to the language games of which they are a part. But one might ask whether or not one should use these terms at all. Aside from historians of science, no one uses the term “phlogiston” anymore. Very few use the term “spinster.” “Heretic,” along with Milton’s “golden-tressed sun” running “his” daily course, is gradually losing traction in Western societies, while “homophobe” (along with the sun-orbiting Earth) is gaining ground. In order to make sense of and justify these shifts from one vocabulary to another, one might think that we need some norms that transcend particular language-games, norms for choosing which vocabulary is best all things considered, which best maps onto the world as it really is. Rorty resists this move: “the world does not tell us which language games to play . . . [C]hange of this magnitude does not result from applying criteria . . . Philosophy, as a discipline, makes itself ridiculous when it steps forward at such junctures and says that it will find neutral ground on which to adjudicate the issue” (Rorty 1988:5, 51). Instead, he argues that one’s choice of vocabulary, too, is ultimately local, though in a different sense. We should, he suggests, “treat alternate vocabularies . . . like alternative tools” (Rorty 1988:11). In such a context,

We will not be inclined to ask questions like ‘what is the place of consciousness in a world of molecules?’ . . . ‘what is the place of

value in a world of fact?' . . . 'What is the relation between the solid table of common sense and the unsolid table of microphysics?' . . . 'Merely philosophical' questions, like [the] question about the relation between the two tables, are attempts to stir up a fictitious theoretical quarrel between vocabularies which have proved capable of peaceful coexistence. (Rorty 1988:10-11)

Choice of vocabularies (or language-games, or paradigms) is local in the immediate sense that one chooses one's vocabulary *for a purpose*. For heightening religious devotion, Milton's description of the sun is useful. For determining when to go to the beach to enjoy the sunset, one should turn to the weather channel. And for determining when (and at what speed) to launch a satellite, one should use the scientific description. This emphasis on the *utility* of vocabularies, and thus for the basic structures in terms of which we think about our world and our lives, is what makes Rorty's philosophy a sort of "pragmatism." What we say and think should be "practical;" it should help us get things done.

But Rorty's norms are "local" in another sense as well, since he rejects the idea of a practical or moral master vocabulary just as much as he rejects the notion of an epistemic master vocabulary. If we could determine a priori what "should be done," then we might be able to determine what to say and think based on some categorical imperative of action. If we could know that religious devotion is more important than satellites, we would know we should prefer Milton to astrophysics. Or if we could know that religious devotion is dangerous superstition, we could justifiably send Milton to the scrap-heap (or, perhaps better, put him to use in new ways). But for Rorty, one determines what to *do* based on norms intrinsic to particular language-games, and, more generally, intrinsic to the particular forms of life that provide social context for those language-games.

I do not think there are any plain moral facts out there in the world, nor any truths independent of language, nor any neutral ground on which to stand and argue that either torture or kindness are preferable to the other. (Rorty 1988:173)

Just as one cannot stand above every different paradigm in physics and ask which is best "all things considered," so one cannot stand above every different conception of human life and ask what one ought to do "all things considered" (see Rorty 1988:50-3, 85-7).

One might think, of course, that one of the things we should be doing with at least some of our vocabularies is "getting the world right."

In the last chapter, I pointed out that even if Heidegger is right (which he certainly is) that we are not always engaging with the world as reflective objective knowers, there is still room (as Kant insists) for thinking that knowing the objective world is at least *one* worthwhile epistemic project (and perhaps even a particularly important one). Rorty rejects this suggestion and so normativity ends up being “local” both in the sense that particular norms are always intrinsic to particular language-games and in the deeper sense that the norms governing one’s choice of language-game are tied to locally-relevant conceptions of what tools are needed to live a good life. The result is what Rorty calls a “cheerful ethnocentrism,” according to which one can happily accept that “all reasoning is tradition-bound” and not try to find some “neutral ground” beyond one’s own tradition (Brandom 2000:20).

What counts as being a decent human being is relative to historical circumstance, a matter of transient consensus about what attitudes are normal and what practices are just or unjust. (Rorty 1988:189).

One might think that the recognition of one’s values as being historically contingent would undermine one’s commitment to them, but Rorty rejects this suggestion: “belief can still regulate action, can still be thought worth dying for, among people quite aware that this belief is caused by nothing deeper than contingent historical circumstance” (Rorty 1988:189). Rorty’s pragmatism aims to cure us of precisely this insistence upon indubitable and final justification for one’s beliefs. Once we stop using language of objectivity and “absolute” truth, we need not take the contingency of our beliefs as undermining them. One’s commitment to any particular value will depend upon one’s particular projects, the value of which will, in turn, depend upon other values. As Rorty puts it, “*We* have to start from where *we* are” (Rorty 1988:198). Thus ethnocentrism need not be a “relativist” undermining of all commitment to our beliefs and values, but can be a “cheerful” acceptance of them.

Even or perhaps especially when cheerful, however, ethnocentrism can seem like a vice to avoid. For one thing, “cheerful ethnocentrism” seems to suggest a stale conservatism of beliefs and values. If norms are intrinsic to language-games and related forms of life, it might seem that these norms are never subject to really whole-hearted revision. If one could justify this stasis on the basis of truth, on the basis that one has finally figured out what the world is *really* like and what a good human life *really* consists in, then it might not be so bad. But Rorty’s cheerful ethnocentrism can offer no such assurance. So ethnocentrism seems unjustifiably stagnant and thus not very cheerful. Moreover,

“ethnocentrism” can lead to ignorance about and intolerance towards other “ethnic” groups; at worst, without non-local bases for reasoned agreement with others, ethnocentrism in our pluralist and interconnected world even seems to invite imperialism and oppression.

One of the great merits of Rorty’s work is that he takes on these concerns and emphasizes the importance of both a creativity that outstrips the norms of one’s language-games (Rorty calls this “irony”) and a tolerance that seeks to understand, appreciate, and protect others’ perspectives (Rorty’s “liberalism”). Rorty’s emphasis on “irony” manifests itself particularly poignantly in his rejection of philosophical intuitionism. One might expect a cheerful ethnocentrist to embrace the prevailing intuitions of his “ethnos,” but Rorty does no such thing:

On the view that I am suggesting, the claim that an “adequate” philosophical doctrine must make room for our intuitions is a reactionary slogan, one which begs the question at hand . . . What is described as such a consciousness is simply a disposition to use the language of our ancestors . . . Unless we suffer from what Derrida calls “Heideggerian nostalgia,” we shall not think of our intuitions as more than platitudes, more than the habitual use of a certain repertoire of terms, more than old tools which as yet have no replacements. (Rorty 1988:22)

Admitting that one must take one’s intuitions, *for now*, as “tools that have no replacements,” Rorty rejects both Heideggerian nostalgia that looks back for originary insights and Moorean conservatism that takes present intuitions as eternal truths. Instead, he opts for a Nietzschean optimism that hopes for new intuitions in a better and brighter future that we can make for ourselves. For Rorty, this new future is, in part, social, in that new terms and visions can catch hold and become shared intuitions of our descendents. But Rorty is also passionately committed to a personal ideal of irony, one represented by poets who transcend norms, terms, and tools (linguistic and otherwise) of the society in which they live, affirming their unique selves (Rorty reads Derrida as a poet of this sort). The ironist in this sense is the existentialist par excellence, the one who not only recognizes her freedom from any norms but actualizes this freedom through creatively reinterpreting and thereby rising above currently dominant “intuitions.” Rather than stale conservatism, cheerful ethnocentrism enriched by irony leads to radical and playful revisions of dominant norms and intuitions in the service of personal authenticity.

Irony on its own, however, does little to alleviate the ethical concern that associates ethnocentrism with intolerance and imperialism.

A willingness to violate, transcend, and reinterpret the values of one's culture need not imply any greater tolerance for the values of others and often leads to the opposite. Rorty is well aware of "how our attempts at [authenticity], our private obsessions with the achievement of a certain sort of perfection, may make us oblivious to the pain and humiliation we are causing" (Rorty 1988:141). Unlike Kant, Rorty does not think that there is anything that connects "our duties to self and duties to others" (Rorty 1988:141), no a priori principle that shows that the way to *really* be creative is to allow others the space to be creative in turn, no argument from the conditions of possibility of irony to the necessity of respect for others. But, Rorty insists, one's commitment to respecting others need be none the less for the lack of an a priori principle (see Rorty 1988:189). Rorty is a *liberal* ironist, and this liberalism brings with it two important commitments that mitigate irony's free-for-all creativity. First, "liberals are the people who think that cruelty is the worst thing we do" (Rorty 1988:xv).³¹⁴ As a liberal, Rorty is sensitive to – and aims to be increasingly sensitive to – acts and attitudes that are cruel towards others. And Rorty goes further than merely *supplementing* ironic ethnocentrism with liberalism. He insists that ironic creativity *rather* than transcendental justification is what liberals really need. For one committed to being less cruel – and Rorty sees himself and his audience as being so committed – what is needed is not an assurance that we are right to avoid cruelty, but greater sensitivity to the variety of ways of being cruel. And cultivating this sensitivity likely requires revisions of perspective. When Foucault shows how supposedly "humane" prison conditions are actually dehumanizing forms of manipulation and control, when Nabokov shows the cruelty of pursuing personal fulfillment against social norms, and when Levinas shows how systems of "justice" impose de-facing categories on others, they all draw attention to forms of cruelty that can be obscured by "intuitions." In that sense, Rorty suggests, ironic overcoming of philosophical commitments to "objectivity" and "justification" actually contributes to better liberalism.

Rorty's liberalism includes a further element that alleviates concerns with oppressive and complacent ethnocentrism. While rejecting "objectivity" and Kantian "universality," Rorty insists upon an ironic-liberal ideal that he calls "solidarity." Solidarity includes "the imaginative ability to see strange people as fellow sufferers" (Rorty 1988:xvi), so is an important part of avoiding cruelty. But solidarity also gives new meaning to "objectivity":

³¹⁴ Of course, "There is no neutral, non-circular way to defend the liberal's claim that cruelty is the worst thing we do" (Rorty 1988:196).

For pragmatists, the desire for objectivity is not the desire to escape the limitations of one's community, but simply the desire for as much intersubjective agreement as possible, to extend the reference of "us" as far as we can. (Rorty 1990:23)

Rather than trying to conform beliefs to the "real world," liberal ironists aim for an epistemic, social, and political community that includes more and more diversity. The negative connotations of "ethnocentrism" are avoided by this desire to justify beliefs and actions to as wide an audience as possible while avoiding cruel attempts to destroy or undermine that audience's diversity:

What takes the curse off [our] ethnocentrism is . . . that it is the ethnocentrism of a "we" ("we liberals") which is dedicated to enlarging itself, to creating and ever larger and more variegated *ethnos*. It is the "we" of a people who have been brought up to distrust ethnocentrism. (Rorty 1988:198)

Rorty is not an ethnocentrist who excludes other people and their points of view out of hand, but one who seeks to *include* them in an ever-broadening, peaceful, and multifaceted perspective.

There is something appealing about Rorty's cheerful ethnocentrism.³¹⁵ Rorty is correct that, from-within, we have no choice but to take ourselves as we are. Even when we seek to incorporate other perspectives, we take them seriously by, in some way, making them our own. And any concerns that raise problems for my presuppositions raise those problems only by appealing to some other presupposition. Even concerns about ethnocentrism, the sense that diversity might undermine objectivity, and the desire for a neutral standpoint are all issues arising from *our* perspective. Ethnocentrism is right as far as it goes; we always wear our own shoes. And this *might* be as much as we can say. If so, then I share Rorty's conviction that we need as much liberalism, compassion, and tolerance as we can reasonably muster. But Rorty's positive arguments for local norms are much stronger than his negative arguments against any possibility of developing norms that are *less* local and perhaps even universal. There may still be hope for normativity that is more robust, less ethnocentric, and more "objective," than Rorty's.

And we – at least many of us, at least some of the time – *need* something more than what Rorty provides. Rorty's liberal irony provides

³¹⁵ Rorty's is not the only cheerful ethnocentrism, nor the only appealing one. For two alternative philosophical approaches that also fit into the cheerfully-ethnocentric mold, see Blackburn 1998 and Timmons 1999 (e.g. p. 218).

little guidance one genuinely concerned about what to believe or how to act. Even if already engaged in language-games with their own internal norms, I can wonder about whether I *should* be thinking and acting in the light of these norms. Rorty writes approvingly of Freud, who “does not ask us to choose” between the “merely moral man” and the “poet,” noting that Freud “does not think we have a faculty which can make such choices” (Rorty 1988:35). But from-within, we often *must* make such choices. And even if I have already accepted that “cruelty is the worst thing one can do” and already endorse a personal ethic of creativity and self-improvement, I not only face potential conflicts between these ideals but can also raise questions about the legitimacy of these ideals. Rorty’s way of dealing with this existential possibility is to focus on particulars, to look at specific and local reasons for preferring one ideal over another. His solution is simply to “get over it,” to stop reflecting too much. There is no need to ask questions about whether what I ultimately care about most is *worth* caring about; and Rorty proposes revising philosophical and even everyday language to make such questions increasingly difficult to ask.³¹⁶

For Kant, however, human beings from-within are always able, at least in principle, to ask such questions, and “enlightenment” or “autonomy” precisely consists in refusing to let these questions be silenced, refusing to let oneself simply absorb the categories, norms, and values of others. Even if (as Foucault and Rorty suggest) this ideal of autonomy is a historically-local “Enlightenment” ideal, it is an ideal that “we” share. Rorty-the-ironist agrees with his Romantic and Nietzschean predecessors that the autonomy worth caring about does not depend upon universal standards. And he offers examples, in his own writing and those of his ironist heroes, of creativity arising from being set free from these standards. But this absence of standards and Rorty’s cheerful ethnocentrism both leave one open to the questioning, which Kant claims is inevitable and which Rorty has not yet managed to cure, of whether one’s ironic play and liberal hope is really worth it, after all. Rorty fails to fully answer the normative question in the way that it shows up, or at least can show up, to human beings.

Moreover, the world today is increasingly pluralist in increasingly deep ways. Rorty is right that one of the greatest merits of “our” liberal societies is our desire to be inclusive, to be fair to those marginalized by “us.” But he underestimates the extent to which this ideal, even if it is

³¹⁶ Rorty might add that the ability to ask “but should I really care so much about cruelty?” is an ability born of language games of justification in terms of an external world, absolute norms, and, ultimately, from language-games in which human beings subordinate themselves to a God who stands above and apart.

only “our” ideal, commits us to attempt non-ethnocentric ways of arbitrating between diverse forms of life. Rorty often mistakenly portrays this need for arbitration as fundamentally *political*. There is a political problem of how to respect those whose values differ fundamentally from one’s own, a problem dealt with, in large part, through keeping individuals and communities as free from political interference as possible (that is, by minimizing the extent to which we need to agree with one another in order to live well together). But there is also a more *personal* problem, one that arises for modern liberal individuals from-within, a problem of how to decide what *I personally* should believe or do in the light of my own contact with different communities who hold incompatible visions of what the world is like and how to live within it. In the modern world, one can no longer simply assume the vocabulary of one’s birth. Before one can be a cheerful ethnocentrist, one needs to decide which ethos (or better, ethoi) to make one’s own.

III. Local Normativity, Narrative Unity, and Tradition(s)

Given that most of us recognize, in the world in which we live, multiple incompatible possible sets of norms for governing how to think and act, we require some way of arbitrating between them. A brute intuitionist approach posits norms that emerge on reflection as obviously or intuitively right, in the light of which one can evaluate other, more local norms. Cheerful ethnocentrism lets us cheerfully accept the evaluative structures of our present tradition without suggesting that anyone in any culture would be able to rationally accept these. Both approaches risk underplaying not only the diversity of norms, but the extent to which, for many people, this diversity has become (or constantly threatens to become) internalized. In the final two sections, we will look at a pair of broadly Kantian approaches that seek to construct a universalist order that can have priority over the intuitions, values, and language-games of particular cultures. Before turning to those universalist, Kantian approaches, however, this chapter looks at a richer and more well-worked out pluralist and non-universalist approach to normativity. The central idea behind this approach is that the primary structure of human deliberation about what to think and believe is not logical and foundationalist, nor even strictly coherentist, but *narrative* in structure.³¹⁷ One of the most important philosophers who uses

³¹⁷ In some respects, one can see this approach as another variation of cheerful ethnocentrism, one that develops a specific account of the normative structure of decision-making about how to live one’s life. Rorty even adopts a similar view in describing how one chooses vocabularies for one’s personal life:

“narrative unity” as a framework for thinking about what the norm-choice is Alasdair MacIntyre.³¹⁸ MacIntyre’s account shares many of Rorty’s central claims about the local character of normativity. But MacIntyre helps reveal an underlying (and even “universal”) normativity in the notion of a “narrative unity” ties together appeals to local norms. Moreover, MacIntyre situates individual narratives in the context of communal narratives – traditions – and develops an account of “the rationality of traditions” that, without providing a “master vocabulary” or “view from nowhere,” still provides a general structure within which one can justify preferring one tradition (and hence one language-game) over another.³¹⁹ MacIntyre thereby provides more general normative claims that the *mere* cheerful ethnocentrism of Rorty, without requiring absolute and strictly universal normative principles that rise above all contingent and particular human communities.

MacIntyre insists upon the narrative structure of individual choices: “Narrative history . . . [is] the basic and essential genre for the characterization of human actions” (MacIntyre 1984:208), so one can make an action, statement, belief, or choice “intelligible by finding its place in a narrative” (MacIntyre 1984:210). Narrative structure does not merely make sense of *others*, but is also essential for the intelligibility of thought and choice “to agents themselves” (MacIntyre 1984:208):

[A]ction itself has a basically historical character. It is because we all live out narratives in our lives and because we understand our own lives in terms of the narratives we live out that the form of narrative is appropriate for understanding the actions of others. (MacIntyre 1984:212)

One decides which language-games to play, which purposes to pursue, what kind of beliefs to hold or actions to take, in terms of an overall life-narrative. “To ask, ‘what is the good for me?’ is to ask how best I might live out that unity [“of a narrative embodied in a single life”] and bring it completion. To ask, ‘What is the good for man?’ is to ask what all answers to the former question must have in common” (MacIntyre 1984: 218-9).

There are two principle ways in which human beings try, by placing their lives in a larger context, to give sense to those lives. The first [which Rorty here endorses] is by telling the story of their contribution to a community. This community may be the actual historical one in which they live, or another one, distant in time or place, or a quite imaginary one . . .” (Rorty 1990: 21)

³¹⁸ MacIntyre’s appeal to the importance of narrative and tradition is part of a more comprehensive critique of an “Enlightenment Project” that includes Kant, a critique that I do not discuss here. See MacIntyre 1984:36-61.

³¹⁹ MacIntyre also proposes certain general *virtues* necessary for any successful narrative unity of life or rational progress of a tradition.

This approach effectively captures much day-to-day experience of the normative character of choice. As existentialists recognized, decisions about life involve incorporating one's past into a projected future; they have a narrative structure. One looks back, seeing roles and responsibilities one has taken on, and thinks about how one's present and future actions can "complete" one's life so far. This largely explains the sense of being "trapped" in a "dead-end" career, or relationship, or habit. It can become difficult to envision a coherent story that goes from what one has been and where one finds oneself towards any complete life that one would find worth living. Without narrative progress, one's life is at a "dead-end."

This account of the narrative structure of life brings out two features that, for MacIntyre, are essential to making sense of normativity in narrative terms. First, a "crucial characteristic of all lived narratives" is "a certain teleological [or goal-oriented] character" (MacIntyre 1984:215). Second, and more importantly, one's life-narrative is interconnected with one's community. Not only does one depend upon that community for support in living one's life, but "we approach our own circumstances as bearers of a particular social identity" (MacIntyre 1984: 220).

I can only answer the question, "What am I to do?" if I can answer the prior question, "Of what story or stories do I find myself a part?" We enter human society, that is, with one or more imputed characters – roles into which we have been drafted – and we have to learn what they are in order to be able to understand how others respond to us and how our responses to them are apt to be construed. (MacIntyre 1984:216)

Neither the child of immigrant parents who have worked in order to send her to college nor the fourth generation child of a long line of Marines who has been groomed for a career in the military can construct a personal narrative according to which she just falls into plumbing as a career. If she chooses to be a plumber, it will be a form of rejecting or radically reinterpreting the role into which her parents' goals and sacrifices have thrust her. We find ourselves, first as children (and siblings), then as students, friends, spouses, workers of particular kinds, and so on, in a host of social roles. These social roles are defined by one's community, and while one can appropriate, revise, or reject these roles, one cannot construct one's personal narrative independent of them. In that sense, "We are never more (and sometimes less) than the co-authors of our own narratives" (MacIntyre 1984:213). "What ought I to do?" quickly becomes, even from-within, "How should balance my family

responsibilities (as spouse/parent) with my teaching responsibilities (as professor)?”

While individuals choose in the light of stories constructed around roles within a community, communities too have stories, so “the story of my life is always embedded in the story of those communities from which I derive my identity . . . I find myself part of a history and that is generally to say, whether I like it or not, whether I recognize it or not, one of the bearers of a tradition” (MacIntyre 1984:221). Just as one sees choices about what to do with one’s life teleologically, that is, as being about how to bring to completion the life that one has lived thus far, one sees these choices also as choices about how to be part of the progress of one’s community towards goods that can make sense in terms of that community’s narrative:

All reasoning takes place within the context of some traditional mode of thought, transcending through criticism and invention the limitations of what had hitherto been reasoned in that tradition . . . A living tradition then is an historically extended, socially embodied argument, and an argument precisely in part about the goods which constitute the tradition. (MacIntyre 1984:222)

So far, this sounds like Rortian ethnocentrism, albeit with a narrative twist. And MacIntyre, like Rorty, vehemently rejects “some neutral tradition-independent standard of a rationally justifiable kind to which we may appeal” in “decid[ing] between rival and mutually exclusive claims . . . [T]here is no such neutral ground . . . [T]here is instead only the practical-rationality-of-this-or-that-tradition and the justice-of-this-or-that-tradition” (MacIntyre 1988:329, 346, see too 166).

But MacIntyre differs from Rorty in two fundamental respects. One, shared by many cheerful ethnocentrists who reject Rorty’s extreme pragmatism, is that MacIntyre sees an important role for “truth” within tradition-guided enquiry. Where Rorty sees “the idea of Truth as something to be pursued for its own sake” as “the central theme” of a *particular* tradition beginning with Plato and, he hopes, ending with himself (Rorty 1990:21), MacIntyre sees the pursuit of Truth as intrinsic to *any* tradition of enquiry: “implicit in the rationality of [tradition-constituted] enquiry there is indeed a conception of final truth” (MacIntyre 1988:360). Thus “every tradition, whether it recognizes the fact or not, confronts the possibility that at some future time it will fall into a state of epistemological crisis,” such that “that particular tradition’s claims to truth can . . . no longer be sustained” (MacIntyre 1988:364).

This leads to a second difference from Rorty: although MacIntyre rejects “neutral standards” for evaluating traditions, he develops a theory of traditions’ rationality that allows for a supra-traditional evaluative framework that can “answer relativism” in a way Rorty is uninterested in doing:

The grounds for an answer to relativism and perspectivism are to be found, not in any theory of rationality as yet explicitly articulated and advanced within one or more of the traditions with which we have been concerned, but rather with a theory embodied in and presupposed by their practices of enquiry, yet never fully spelled out, although adumbrations of it, or of parts of it, are certainly to be found in various writers . . . The rationality of a tradition-constituted and tradition-constitutive enquiry is in key and essential part a matter of the kind of progress which it makes. (MacIntyre 1988: 354)

There are two senses of the “rationality of traditions” at play here. *Within* each tradition are standards of rationality that must be met: “traditions possess measures to evaluate their own progress or lack of it, even if such measures necessarily are framed in terms of and presuppose the truth of those central theses to which the tradition gives allegiance” (MacIntyre 1988:167). But there are also meta-standards operative in and between traditions , meta-standards that can never be fully articulated because any articulation is in terms of a particular tradition, but that can nonetheless be glimpsed, especially in the context of historical evolutions of and confrontations between traditions. MacIntyre focuses on what he calls, recalling Hegel, a “dialectical” rationality at play within and between traditions:

The kind of justification that they [first principles within a tradition] receive is at once dialectical and historical. They are justified insofar as in the history of this tradition they have, by surviving the process of dialectical questioning, vindicated themselves as superior to their historical predecessors. (MacIntyre 1988:360)

In laying out how this vindication might go, MacIntyre explains,

[P]rotagonists of each tradition . . . ask whether the alternative and rival tradition may not be able to provide resources to characterize and to explain the failing and defects of their own tradition more adequately than they, using the resources of that tradition, have been able to do. (1988: 167)

Thus, for example, MacIntyre explains why Aquinas’s philosophy rightly deserved to win out over its predecessors:

What justifies his representation of the order of things over against its Averroist, Neoplatonist, and Augustinian rivals is its ability to identify, to explain, and to transcend their limitations and defects, while preserving from them everything that survives dialectical questioning in a way which those rivals are unable from their philosophical resources to provide any counterpart. (MacIntyre 1988: 172, see too 175-6)

With this dialectical account of tradition-development, MacIntyre offers a general approach to diversity that steers clear of mere ethnocentrism without articulating specific principles of rationality to which all traditions must subscribe.³²⁰

For Kant, MacIntyre's emphasis on narrative unity and even finding one's place in the tradition can be entirely welcome. Especially in the context of his account of radical evil, Kant points out the importance of seeing one's whole life as a narrative of moral progress, and Kant can accommodate the need for one's thoughts and actions to be consistent with while still pushing forward those of the traditions in which one finds oneself. Nonetheless, for Kant, there are several problems with MacIntyre's account. MacIntyre's approach to narrative unity is too vague. Where MacIntyre simply posits narrative unity as an ideal, Kant suggests specific means of bringing life to coherent unity through cultivating a character that acts consistently in accordance with principles and through the unification of these principles into a moral whole. Moreover, MacIntyre's "narrative unity" is insufficiently autonomous, since it allows for a primarily passive relationship to substantive values that one adopts from one's tradition.

With respect to tradition, not only is the dependence upon tradition overly heteronomous, but while MacIntyre's conception of tradition is a valuable descriptive account of why people hold the values – and even the "rationality" – that they do, it is insufficient from-within, when one is trying to decide whether and to what extent one should

³²⁰ Finally, MacIntyre complements his emphases on narrative and on traditions with an emphasis on *virtues*. Among other things, virtues "sustain . . . the form of an individual life in which that individual may seek out his or her good as the good of his or her whole life" and "sustain and strengthen traditions" (MacIntyre 1984: 223). Virtues like "justice . . . truthfulness . . . courage . . . [and] relevant intellectual virtues" (MacIntyre 1984: 223) transcend individual narratives and traditions while still being part of them. Traditions or narratives that fail to value properly justice or truthfulness or courage will not – and, one might even say, *should* not – be sustained. Thus despite starting from a deeply historicist and narrative conception of normativity, MacIntyre finds non-historical and non-narrative norms, virtues that make possible any "history" with normative import, any narrative that can provide unity of life.

endorse the values and “rational” norms of one’s traditions. MacIntyre’s insistence upon “truth” as a telos of traditions of enquiry and his supra-traditional evaluative framework improve on Rorty’s cheerful ethnocentrism, but MacIntyre’s account of this framework is too amorphous and arbitrary to sustain the from-within needs of 21st century thinkers and actors who find themselves only loosely tied even to their own traditions. MacIntyre’s proposal is too amorphous in that it recommends merely “the process of dialectical questioning,” offering, within that framework, historical case studies but little that can guide today’s pressing choices between traditions. And the proposal is too arbitrary in that its Hegelian historical lineage is all-too-obvious, opening the question of why one should take this quasi-Hegelian conception of rationality as *the* definitive supra-traditional one.

MacIntyre’s approach is strongest where it is most Kantian. Although MacIntyre specifically contrasts his defense of “virtues” to “the knowledge of a set of . . . maxims which may provide our practical inferences with major premises” (MacIntyre 1984: 223), he defends virtues as universally required because they provide necessary conditions for the possibilities of successfully unifying life and of any tradition’s rational development. They are general conditions – albeit qualities of character rather than moral principles – of the possibility of normativity.

IV. Habermas and Communicative Rationality

While contemporary neo-Kantians incorporate many insights of the approaches discussed in this and previous chapters, they retain Kant’s emphases on universality, autonomy, and the unconditionality of the moral law. The rest of this chapter focuses on two of the most important contemporary neoKantian (moral) philosophers, starting in this section with Jürgen Habermas and turning in the next to Christine Korsgaard.

Jürgen Habermas is a leading defender of a broadly Kantian approach to normativity, especially in moral and political contexts. Habermas offers a Kantian version of the “rationality of traditions” developed by MacIntyre. Like MacIntyre and Rorty, Habermas recognizes rationality embedded in particular cultural traditions, and he admits that many decisions about how best to live take place in the context of roles and norms inherited from our culture. But, like MacIntyre and against Rorty, Habermas also insists that particular traditions, at least insofar as these are traditions of enquiry, embody a commitment to “truth” that

appeals to something supratraditional and functions as a regulative ideal (see especially Habermas 1993:96-105). Moreover, like MacIntyre, Habermas insists that pursuing truth in the context of conflicting traditions involves a “dialectical and historical” process (MacIntyre 1988: 360, cf. e.g. Habermas 1990:4-5, 87). But unlike MacIntyre (and Rorty), Habermas argues that this dialectical process involves certain universal and necessary norms of interaction that can be articulated and defended.³²¹ Habermas thereby aims to reconstruct Kantian moral philosophy as a “discourse ethics” or “theory of communicative action” that articulates universal norms that govern and arise within the concrete efforts of members of different traditions to come to agreement.

Habermas’ argument can be seen as extending MacIntyre’s appeal to the rationality of traditions, but Habermas describes it as a Kantian (re)appropriation of Hegel’s insight that normativity and rationality are fundamentally public and social. In Habermas’s hands, Hegel’s social rationality gives rise to an “intersubjective interpretation of the categorical imperative” (Habermas 1993:1).³²² On this view,

Controversial questions of normative validity can be thematized only from the first person plural perspective, that is, . . . “by us”: for normative validity claims are contingent on “our” recognition. (Habermas 1993:49)

In that case, “it becomes imperative to submit all norms to a public, discursive, generalization test that necessitates reciprocal perspective taking” (Habermas 1993:64). For Habermas, the need to develop norms within public discourse and subject them to such discourse provides for “grounding the moral principle in the normative content of our practice of argumentation” (Habermas 1993:30) and thereby for “the distinctive idea of an ethics of discourse”:

(D) Only those norms can claim to be valid that meet (or could meet) with the approval of all affected in their capacity *as participants in a rational discourse* (Habermas 1990:66, see Habermas 1993:31-2, 50)

³²¹ MacIntyre hints of this in his tacit endorsement of Aquinas’ appeal to natural law, as when he suggests, “the natural law is discovered not only as one of the primary objects of practical enquiry but as the presupposition of any effective practical enquiry” (MacIntyre 1988:180). In the main, however, MacIntyre eschews the sorts of universal conditions of legitimate dialogue that Habermas and Aquinas seek to articulate.

³²² Even Rorty admits that “[t]here are no constraints on inquiry save conversational ones – . . . only those retail constraints provided by the remarks of our fellow-inquirers” (Rorty 1982:165). Habermas extends this concession of conversational constraints into a wholesale normative theory.

This procedural principle for norm-validation is a sort of proto-correlate to the categorical imperative. In itself, it has no content, but discourse ethics articulates further principles constitutive of genuinely rational discourse. A rational discourse, or a truly “communicative” interaction, aims for “truth . . . rightness . . . truthfulness” and ultimately mutual, rational motivation, as opposed to mere “strategic action” where “one actor seeks [merely] to *influence* the behavior of another,” say through manipulation, deceit, or coercion. Throughout his writings, Habermas articulates norms – such as “openness, equal rights, truthfulness and the absence of coercion” – constitutive of communicative action (Habermas 1993:32). Rather than a moral theory laying out what individuals ought to choose, Habermas’ ethics of discourse elucidates standards governing how communities seeking norms together ought to speak with one another. From these standards, discourse ethics provides an ethics of discourse in further sense, in the various provisional and fallible normative claims that emerge *from* ethically-conducted discourses.

Throughout, Habermas insists on “actual participation [in] . . . real discourse” (Habermas 1990:67-8, cf. Habermas 1993:51) to discern practical moral norms for action. One cannot *simply* analyze conditions of possibility of communicative action as a “ideal . . . isolated subject” (Habermas 1993:48): “no norm . . . can be justified . . . privately in the solitary monologue of the soul with itself” (Habermas 1993:64, see too Habermas 1990:65-7). This gives rise to an important Habermasian “fallibilism” (Habermas 1993:38). Because “the social world . . . is *historical*” (Habermas 1993:39), norms that seem justified in particular historico-social contexts may no longer be justified in later ones. In terms of the principle (D) above, not only does the sphere of “those affected” by a given norm change over time, but one might think a given norm “could meet” with approval by all affected only to find, in the course of *actual discourse*, that it cannot. Without succumbing to quasi-relativist historicism, Habermas allows that the communicative process whereby human beings form universal norms is subject to the imperfections of real, historical, dialogues.

Still, Habermas defends a Kantian approach to normativity, but modifies Kant in response to recent insights about human beings. The position is recognizably Kantian in emphasis on universality; Habermas’ ideal of deliberative consensus is a more down-to-earth instantiation of Kant’s imaginary “realm of ends.” But Habermas rejects what he sees as Kant’s overly individualistic approach to norm-formation, replacing an isolated individual assessing beliefs and choices with a *community* discoursing about them. Thereby Habermas, more than Kant, recognizes

the important cultural and historical dimensions of cognition and moral reflection.

Habermas further alters Kant's position by blurring boundaries between transcendental, empirical, and pragmatic anthropology and thereby developing a more naturalist conception of norms and a more value-loaded empirical psychology than Kant's.³²³ With respect to the former, Habermas engages in "the task of 'situating reason'" by "detranscendentalizing" both "the subject of knowledge" (Habermas 2008:25) and the "acting subject," (Habermas 2008:40). Both are "dislodged from the kingdom of intelligible beings into the linguistically articulated lifeworld of socialized subjects" (Habermas 2008:40, cf. 30, 34).³²⁴ Actual communication in a real world of embodied and socialized human beings gives rise to epistemic and moral norms and thereby to the objective world within which human beings find themselves and the moral demands to which they are subject. What for Kant were a priori transcendental conditions of possibility of cognition and volition become, for Habermas, standards that emerge from concrete acts of communication. There remain, of course, correlates to Kant's transcendental ideas. Habermas discusses how, for example, "the distinction between truth and rational acceptability replaces the difference between 'things in themselves' and appearances" (Habermas 2008:34) or how "counterfactual assumptions" about not-yet-realized communicative situations play a regulative role that balances the natural account of the emergence of moral norms. But even the "must" of the constitutive standards of communicative action itself has, for Habermas,

a Wittgensteinian rather than a Kantian character[, t]hat is, . . . not the transcendental meaning of universal, necessary, and noumenal conditions of possibility . . . but the grammatical meaning of an 'unavoidability' stemming from the conceptual connections of learned – but for us inescapable – rule-governed behavior. (Habermas 2008:27)

For Habermas, norms arise not from some noumenal, first-person-singular realm of thought or choice but from the this-worldly, social, first-and-second-person-plural realm of communicative action.

³²³ The naturalism on which I focus in this paragraph is an even "softer" naturalism than the "soft naturalism" that Habermas endorses in Habermas 2008 (especially chapter 6). My emphasis is how Habermas makes norms this-worldly and thus natural as opposed to "noumenal".

³²⁴ This also provides Habermas an opportunity to emphasize the important role of "second-person" perspectives on others in discourse, appropriating some of the insights of Levinas (see Habermas 2008:40, 1993:15).

While rooting transcendental anthropology in the empirical world, Habermas also insists that empirical investigations of human beings cannot be isolated from moral norms. Habermas turns to the cognitive psychology of Lawrence Kohlberg for an example of an “empirical theory” that “is dependent on the philosophical theories it employs” and that therefore provides “indirect testing” for the (broadly Kantian) normative theories underlying it (Habermas 1993:115, 1990:117). Kohlberg’s research program involves positing stages of moral (and cognitive) development, where “universal ethical principles” mark the highest stage. For Habermas, because of the interdependence between normative assumptions about what constitutes a higher stage of development and the empirical claims to which Kohlberg’s model gives rise, the success of Kohlberg’s research program (partly) justifies a (more-or-less) Kantian conception of moral ideals.³²⁵ This interweaving of normative and empirical is similar to Kant’s pragmatic anthropology in seeking empirical work relevant to achieving moral and cognitive ideals, but Habermas’s reading of Kohlberg makes the empirical and transcendental interdependent rather than complementary. Thus empirical research can actually *challenge* transcendental ideals, rather than merely showing strategies for better achieving them.

Another important difference between Habermas and Kant relates to the criticism – shared by Hegel, MacIntyre, and existentialists – that Kant’s dichotomy between morality and self-love fails to recognize the important role that *non*-moral life-choices play in human self-conceptions. In response to these criticisms, Habermas adds Hegel’s notion of the ethical to Kant’s dichotomy between morality and prudence³²⁶:

The question ‘what should I do?’ takes on a pragmatic, an ethical, or a moral meaning depending upon how it is conceived . . . The rational consideration of an important value decision that affects the whole

³²⁵ The converse also applies. Thus Carol Gilligan’s criticism of Kohlberg’s developmental model (e.g. Gilligan 1984) would challenge the fruitfulness of Kohlberg’s empirical research program *and* its broadly Kantian underpinnings, a point not missed on feminist philosophers who use Gilligan to argue against Kant’s moral theory (e.g. Noddings 1984).

³²⁶ In fact, Habermas goes even further than this:

Kant commits himself to the technical-practical and moral-practical reasons. Communicative action draws on a broader spectrum of reasons: epistemic reasons for the truth of statements, ethical orientations for the authenticity of life choices, . . . narrative explanations, cultural standards of value, . . . and so forth. (Habermas 2008:38)

For the purpose of this short discussion, however, I focus on the ethical.

course of one's life is . . . different [from both pragmatic and moral considerations. It] involves hermeneutical clarification of an individual's self-understanding and . . . questions of a . . . not-failed life. The terminus ad quem of a corresponding ethical-existential discourse is advice concerning the correct conduct of life and the realization of a personal life project. Moral judgment . . . is something different. It serves to clarify legitimate behavioral expectations in response to interpersonal conflicts results from the disruption of our orderly coexistence by conflicts of interests. (Habermas 1993:8-9)

As we have seen, moral judgments are extremely important for Habermas, and he preserves Kant's emphasis on their importance. But like existentialists, Habermas also recognizes the important role of decisions that affect whether one's life as a whole fails or succeeds in *non-moral*, existential, terms. Even with respect to these decisions, however, "there is room for discourse because here too the steps in argumentation should . . . be comprehensible in intersubjective terms. The individual attains reflective distance from his own life history only within the horizon of forms of life that he shares with others and that themselves constitute the context for different individual life projects" (Habermas 1993:11). Habermas adds to this existentialist insistence on authenticity a MacIntyrean emphasis on shared forms of life, but he does *not* accept MacIntyre's claim that the moral simply gets *absorbed* into the ethical. Instead, for Habermas, there is an important distinction between *ethical* values, which are always rooted in particular social and cultural traditions but can also be highly individual, and *moral* norms, which "must detach themselves from the concrete contents of the plurality of attitudes . . . that now manifest themselves" to be "grounded solely in an abstract social identity that is henceforth circumscribed only by the status of membership in *some* society, not in this or that particular society" (Habermas 1993:47). Habermas, like Kant but with a greater awareness of the not-merely-pragmatic importance of non-moral choices, advocates a "morally justified pluralism of life projects and life-forms" (Habermas 1993:75). And he adds that as pluralism grows and thereby "the sphere of questions that can be answered rationally from the moral point of view shrinks . . ., finding a solution to these few more sharply focused [moral] questions becomes all the more critical to coexistence" (Habermas 1993:91).

From the standpoint of the developments in our understanding of the human being over the last two centuries, Habermas's communicative ethics has important advantages over Kant's anthropology. Habermas's

moderate naturalism provides a way of jettisoning what many see as needless metaphysical baggage from Kant's philosophy. And in his appropriation of Kohlberg, Habermas provides an important and still broadly Kantian alternative to Kant's pragmatic anthropology for thinking about the relationship between empirical psychology and normativity. His fallibilism provides a way of accommodating historicism without succumbing to it, and his category of the "ethical-existential" even highlights the importance of historically- or culturally-local norms without losing something like Kant's category of moral norms. Finally, Habermas's sociopolitical approach to Kant makes it possible to appropriate Kant's emphasis on morality in the contexts where it seems most important – arbitrating interpersonal and intercultural conflicts – while leaving decisions about how to lead one's life open to a broader range of influences.

But Habermas's differences from Kant bring disadvantages as well. His naturalism compromises the sharp dichotomy between transcendental and empirical anthropology that enabled Kant to maintain his commitment to strong norms of thought and action in the light of naturalist and historicist threats. And the accommodation to existentialist and MacIntyrean approaches to "ethics" opens Habermas to some of the vagueness, indeterminacy, and potential heteronomy of MacIntyre's account of narrative unity. (Where MacIntyre replaces Kantian *morality* with ethics, however, Habermas' ethics only supplements Kant's *hedonism*, which is already vague, indeterminate, and heteronomous.) Perhaps most importantly, Habermas' resistance to allowing the determination of moral norms to take place in individuals, while it has important advantages, also makes his theory less helpful than Kant's for the from-within perspective of individuals seeking to discern what to think and do in non-ideal conditions, where we simply cannot engage in the requisite discussions but must nonetheless decide how to live.

V. Korsgaard and The Sources of Normativity

Like Habermas, Christine Korsgaard is a contemporary Kantian responding to challenges of naturalism, historicism, diversity, existentialism, and alternative normative approaches that have emerged over the past 200 years. Although Korsgaard has written works directly interpreting Kant's philosophy – see especially the essays collected in *Creating the Kingdom of Ends* – her most important work is her Kantian

normative theory, laid out in *The Sources of Normativity*.³²⁷ Throughout, Korsgaard insists that respect for humanity is a “categorical imperative” (e.g. *Sources*, pp. 99-100, 143) and that this imperative normatively binds all human agents. Like Kant, she argues that freedom is a condition of possibility of morality and responsibility and adds that while freedom is, in itself, freedom from determination by natural causes, one can nonetheless affirm that human deeds are both free and causally determined. Finally, she shares Kant’s emphasis on the free but morally-conditioned finitude of human beings as what is, at least for philosophical reflection, what is most important about us.

Still, Korsgaard’s Kantianism differs from Kant’s in several respects, of which three are particularly important: how Korsgaard conceives of human freedom in relation to natural necessity, her defense of morality’s normative status, and the role of particular, contingent “practical identities” in her account. In these areas, Korsgaard’s Kantianism accommodates and responds to the developments discussed in this and the previous three chapters.

With respect to her defense of human freedom, Korsgaard seeks to avoid metaphysically-loaded construals of Kant. Thus Kant’s transcendental idealism is interpreted not in terms of two metaphysically distinct worlds – a phenomenal world of causally determined appearances and a noumenal world within which freedom finds a place – but in terms of two distinct standpoints, a scientific standpoint from which one views the world as a set of causally determined interactions amongst objects and a practical standpoint from which one views oneself as a free agent effecting changes in the world.

Determinism is no threat to freedom . . . The freedom discovered in reflection is not a theoretical property which can also be seen by scientists considering the agent’s deliberations third-personally and from outside. It is from within the deliberative perspective that we see our desires as providing suggestions which we may take or leave. You will say that this means that our freedom is not ‘real’ only if you have defined the ‘real’ as what can be identified by scientists looking at things third-personally and from outside. (Korsgaard 1996b:96)

³²⁷ Korsgaard’s most important recent installment in her normative theory is her *Self-Constitution: Agency, Identity, and Integrity* (Korsgaard 2009), based on her 2002 Locke Lectures at Oxford. While more “up-to-date,” this book is also importantly less Kantian, in that Korsgaard much more explicitly endorses Aristotelian and Platonic moves that *Sources* largely eschews (see especially pp. 1-5) or leaves in the background. That said, while my focus in this section will be on the normative theory of *Sources*, I draw from this later work where appropriate throughout this discussion of Korsgaard.

In itself, this no “departure” from Kant; it articulates the importance of the from-within perspective on which this book has focused. But unlike Kant, Korsgaard tends to see these two standpoints as non-hierarchically arranged; even when mentioning Kant’s “priority of the practical,” Korsgaard more often emphasizes that “neither standpoint is privileged over the other – each has its own territory” (Korsgaard 1996a:173). Korsgaard defends what we called in chapter eight a “neutral” rather than “freedom-first” perspectivism. This difference marks a minor *departure* from Kant, but it represents a sharply metaphysically-deflationary *reading* of Kant, a reading that seeks to maintain as much of Kant’s transcendental idealism as possible while keeping his metaphysics as naturalistic and this-worldly as possible.

The next two points need to be taken together, since Korsgaard’s defense of the importance of morality depends upon her account of practical identities.³²⁸ The most basic difference between the arguments of Korsgaard and Kant can be seen in Korsgaard’s broadly *existentialist* starting point. Her *Self-Constitution* begins with a direct allusion to Sartre: “Human beings are condemned to choice and action” (Korsgaard 2009:1, cf. Sartre 1956:568).

The human mind is . . . essentially reflective . . . And this sets us a problem no other animal has. It is the problem of the normative. For our capacity to turn our attention on to our own mental activities is also a capacity to distance ourselves from them, and to call them into question . . . The reflective mind cannot settle for perception and desire, not just as such. It needs a *reason*. Otherwise, at least as long as it reflects, it cannot commit itself to go forward. (Korsgaard 2006b:92-93).

For Sartre, this problem leads to existential angst, humans’ recognition that they are utterly responsible for their actions. From-within, human beings seek reasons for what we believe and do, which sets us a problem. This problem of reflection, in fact, has haunted each of the last three chapters because, on their own, naturalism, historicism, and even existentialism fail to *address* reflective human beings; they fail to give us *reasons*.

For Kant, humans’ from-within stance is not *merely* existential; there is a “fact of reason” that “forces itself upon us” (5:31) to make us immediately aware of our *moral* responsibility, a *categorical imperative* we ought to obey. Thus *Kant* – at least in the *Critique of Practical Reason* and

³²⁸ This claim is certainly true for Korsgaard 1996b, though the dependence is implicit even in Korsgaard 2009.

later works – sees the normative pull of morality as something *given*, the conditions of possibility for which he seeks (and finds pre-eminently in freedom), but not something that he needs to, or even can, “prove.” But Korsgaard, partly in response to new skepticisms introduced by Nietzsche and his historicist and existentialist heirs, seeks to show that morality’s normative force is built into the merely-reflective existential situation of human beings. She rehabilitates an earlier argument from Kant’s *Groundwork* that emphasizes how “to every rational being having a will we must necessarily lend the idea of freedom also, under which it acts” (4:448), and she argues that Kant is the ultimate conclusion of Sartre’s search for something like an existentialist ethics (see, e.g., Sartre 1947-8/1992).

Korsgaard starts by arguing that human freedom involves searching for normativity, that is, for justified *reasons*. Human beings must choose, and when we choose, we must choose in the light of what we take to be good reasons. The key question is the status of morality within human deliberation; in particular, are we “*justified* in according [special] importance to morality” (Korsgaard 1996b:13). And the answer must *address* one who finds herself in a situation where morality requires significant sacrifices, where one genuinely wonders whether to accord it overwhelming weight in deliberation.³²⁹ Korsgaard then considers classic ways of explaining morality’s normative force, including accounts similar to intuitionism (see her discussion of “realism”) and cheerful ethnocentrism (see her discussion of Hume, Mill, and Williams). In all these cases, reflection does not proceed far enough. When one has an intuition of an action’s rightness, one can ask why one should care so much about that intuition. When one finds oneself – as Rorty does – caring deeply about cruelty, one can ask whether this concern for cruelty is *justified*. And while this question might be dismissed in academic discussions, it cannot be dismissed by one facing real conflicts between other deeply-held interests and her interest in morality.

Like intuitionists and even Rorty, Korsgaard begins by arguing that what it takes for an action to be justified is that it survives careful consideration, what she calls “reflective endorsement” (Korsgaard 1996b:49ff.). Unlike intuitionists, however, Korsgaard denies that such reflection appeals to an abstract realm of real values out there or a set of moral properties inherent in objects. For her existentialist Kantianism, values are *constructed* by human beings; we impose and create values *autonomously*. But unlike Rortyan ironic liberalism, Korsgaard

³²⁹ Here I follow the line of argument in Korsgaard 1996b. In 2009, Korsgaard’s argumentative approach is different, but many key moves end up being the same.

emphasizes the constraints within which values are imposed, the first of which brings her closer to MacIntyre. For Korsgaard, reasons arise from “practical identities,” “description[s] under which you value yourself and find your life to be worth living” (Korsgaard 1996b:101), and these identities can be, and typically are, rooted in contingent historical situations and our local communities. I might value myself as a father, teacher, liberal, wine aficionado, mentor, son, citizen of the United States, husband, human rights advocate, philosopher, mountain-climber, or/and friend. All these identities give life meaning, but only some are so important that I would find my life worthless without them. In various contexts, one or more of these identities provide reasons to act in particular ways, and often, they conflict.

The introduction of practical identities enriches Kant’s picture of obligation in a way that meets needs raised by existentialist, historicist, and communitarian thinkers. Where Kant emphasizes a dichotomy between inclinations and duty, Korsgaard recognizes that humans define themselves in terms of wide varieties of “obligations” that are neither moral nor mere inclination. This ennobles non-moral choices (like existentialists) and moves towards MacIntyre’s emphasis on tradition and sociocultural context. For Korsgaard, cultural and historical contexts do not determine the identities under which we value ourselves. (One’s identity as a child, for instance, cannot be biologically determined, since one can always choose not to *value* oneself under the description of “child.”) But most of our daily decisions involve reasons rooted in *socially, culturally, and historically constructed* identities. When I value myself as “citizen of the United States,” I embrace—perhaps in a mitigated way and/or with desire to change historic conceptions of citizenship—an identity constructed by others over hundreds of years. Korsgaard’s “practical identities” flesh out her broadly Kantian anthropology to incorporate both existentialists’ sense that non-moral choices are intensely important and historicist and traditionalists emphasis on historico-cultural context.

For Korsgaard, however, reflection is not limited to consideration of reasons for action in the context of particular practical identities. One can and should also reflect on one’s practical identities themselves.

The activity of reflection has rules of its own, rules which . . . are constitutive of it.³³⁰ And one of them, perhaps the most essential, is

³³⁰ For more on the notion of “constitutive standards,” see Korsgaard 2009, especially chapter 2. Arguably, *Self-Constitution* arises from Korsgaard’s recognition that the notion of constitutive standards of reflective activity is both central to *The Sources of Normativity* (Korsgaard 1996b) and under-defended there.

the rule that we should never stop reflecting until we have reached a satisfactory answer, one that admits of no further questioning. It is the rule, in Kant's language, that we should seek the unconditioned. (Korsgaard 1996b:257-8)

This insistence upon pushing reflection until we get to an "unconditioned" is Korsgaard's most Kantian move and one contested from various corners. Existentialists despair of ever getting a "satisfactory answer," while historicists, ethnocentrists, and traditionalists like MacIntyre insist that any such answer is always historically-conditioned. But for Korsgaard, no historically-conditioned answer can ever satisfy, since one can always ask, "But why value this particular community or tradition?" And while one might feign cheerful ethnocentrism, such cheerfulness fails to address the demands implicit in reflection. Fortunately, however, against historicist and existentialist pessimisms, it *is* possible to reach an unconditioned, a practical identity about which it no longer makes sense to ask, "But why should I value that?" This identity is one's humanity. Korsgaard's concise argument for this claim is worth quoting in full.

I have argued that human consciousness has a reflective structure that sets us normative problems. It is because of this that we require reasons for action, a conception of the right and the good. To act from such a conception is in turn to have a practical conception of your identity, a conception under which you value yourselves and find your life to be worth living and your actions to be worth undertaking. That conception is normative for you and in certain cases it can obligate you, for if you do not allow yourself to be governed by any conception of your identity then you will have no reason to act or live. So a human being is an animal who needs a practical conception of her own identity, a conception of who she is which is normative for her.

But you are a human being and so if you believe my argument you can now see that that is *your* identity. You are an animal of the sort I have just described. And that is not merely a contingent conception of your identity, which you have constructed or chosen for yourself, or could conceivably reject. It is simply the truth. It is because we are such animals that our practical identities are normative for us, and, once you see this, you must take this more fundamental identity, being such an animal, to be normative as well. You must value your own humanity if you are to value anything at all.

Why? Because now that you see that your need to have a normative conception of yourself comes from your human identity, you can query the importance of that identity. Your humanity requires you to conform to some of your practical identities, and you can question this requirement just as you do any other. Does it really matter whether we act as humanity requires, whether we find some ways of identifying ourselves and stand by them? But in this case, you have no option but to say yes. Since you are human you *must* take something to be normative, that is, some conception of practical identity must be normative for you. If you had no normative conception of your identity, you could have no reasons for action, and because your consciousness is reflective, you could then not act at all. Since you cannot act without reasons and your humanity is the source of your reasons, you must value your humanity if you are to act at all.

It follows from this argument that human beings are valuable. Enlightenment morality is true. (Korsgaard 1996b:122-3³³¹)

On its own, this argument does not establish the legitimacy of Kant's formulation of the moral law because while it gives a reason to value *one's own* humanity, this need not imply valuing the humanity of others. So Korsgaard takes up a point implicit in Kant, but which she draws from Wittgenstein, the point that normative categories are necessarily social. There can be neither private language nor private values. Language makes it possible for others to intrude into our reflection, to give us candidate reasons for action: "If I call out your name, I make you stop in your tracks . . . Now you cannot proceed as you did before . . . you can proceed, but not as you did before" (Korsgaard 1996b:140). Korsgaard's point is a (drastically) tamed-down version of Levinas³³²: other people are capable of obligating us, of requiring us to take them seriously and endorse or reject their needs and sufferings as (candidate) reasons for action. When we honestly reflect about these reasons, we see that we ought to endorse them, at least sometimes, just as we endorse our own desires, at least sometimes: "[I]f you are a law . . . insofar as you are just human, just *someone*, then the humanity of others is also a law . . . In hearing your words as *words*, I acknowledge that you are *someone*," and thus entitled to some degree of respect (Korsgaard 1996b:143).

³³¹ The argument quoted here is unpacked in more detail in *Sources* and in even more detail, though in a slightly less Kantian way, in *Self-Constitution*.

³³² See my discussion of Levinas in chapter ten, pp. xxx-xxx.

There are considerable further nuances and details of Korsgaard's argument, but already it should be clear how Korsgaard is genuinely Kantian, but modifying Kant in the light of 19th and 20th century developments. Her argumentative strategy is at once thinner and more substantive than Kant's. It is thinner in that it starts from mere choice, rather than choice already aware of moral obligations. For Kant a being – though not a “human” being – could be an *agent* without any capacity for being a *moral agent* (see *Religion* 6:26n), so Kant appeals to a “fact of reason” to show how human deliberators are aware of moral demands. Korsgaard, writing after Nietzsche questioned moral universality and after Sartre highlighted humans' universal plight as free beings, reverses Kant's argument to move *from* practical freedom *to* moral obligation. But Korsgaard's argument is also more substantive than Kant's in recognizing that human beings are not pulled merely between a pleasure principle of self-love and a moral principle of duty. Humans are governed by diverse arrays of practical identities. *Moral* obligation –commitment to our identity *as human beings* – is one sort of obligation. And while Korsgaard argues that this sort of obligation is preeminently important, she does so by highlighting its role as underpinning other, contingent practical identities. This move ascribes to choices of particular identities – of friends, lovers, occupations, hobbies, and so on – something of the dignity ascribed to them by existentialists. And it equally importantly gives crucial place in human formation to contingent identities we endorse but do not strictly “choose,” identities that come to us from our cultural and historical contexts.

When combined with her non-metaphysical reading of Kant's transcendental idealism, Korsgaard's philosophy provides a naturalist, historically-sensitive, existential Kantianism. Where Habermas combines these elements in a social way, Korsgaard does so at the level of each individual making choices in his or her own life. The result is a pair of *Kantian* answers that address human agents asking themselves Kant's philosophical but also deeply practical question, “What is the human being?”

Conclusion

Xxx TO BE COMPLETED xxx

Bibliography

- Allen, C., Bekoff, M., & Lauder, G. (eds.) (1998) *Nature's Purposes* Cambridge, MA: MIT Press.
- Allison, H. (1983) *Kant's Transcendental Idealism: An interpretation and defense* New Haven, CT: Yale University Press.
- Allison, H. (1990) *Kant's Theory of Freedom*. Cambridge: Cambridge University Press.
- Allison, H. (2001) "Ethics, Evil, and Anthropology in Kant: Remarks on Allen Wood's *Kant's Ethical Thought*" in *Ethics* 111:594-613.
- Ameriks, K. (1982a, 2d ed. 2000) *Kant's Theory of Mind*, Oxford: Oxford University Press.
- Ameriks, K. (1982b), "Recent Work on Kant's Theoretical Philosophy," in *American Philosophical Quarterly* 19:1-24.
- Ameriks, K. (1982c) "How to Save Kant's Deduction of Taste" *Journal of Value Inquiry* 16: 295-302 (reprinted in Ameriks 2003: 285-307).
- Ameriks, K. (2003) *Interpreting Kant's Critiques*, Oxford: Oxford University Press.
- Ameriks, K. (2000) *Kant and the Fate of Autonomy: Problems in the Appropriation of the Critical Philosophy*, Cambridge: Cambridge University Press.
- Anderson-Gold, S. (2001) *Unnecessary Evil*. Albany, NY: SUNY Press.
- Appiah, A. (2009) "Experimental Moral Psychology" in *Daedalus* 138:92-102.
- Aquila, R. (1983) *Representational Mind: A study in Kant's theory of knowledge*, Bloomington: Indiana University Press.
- Ariew, A., Cummins, R., & Perlman, M. (eds.) (2002) *Functions: New Essays in the Philosophy of Psychology and Biology*.
- Augustine (1961) *Confessions* (trans. R. S. Pine-Coffin), New York: Penguin Classics.

- Baggott, J. (2004) *Beyond Measure: Modern Physics, Philosophy and the Meaning of Quantum Theory*, Oxford: Oxford University Press.
- Baron, M. (1995) *Kantian Ethics (Almost) without Apology*, Ithaca, NY: Cornell University Press.
- Beiser, F. (2002) *German Idealism: The Struggle Against Subjectivism 1781-1801*, Boston: Harvard University Press
- Beiser, F. (1993) *The Fate of Reason: German Philosophy from Kant to Fichte*, Cambridge, MA: Harvard University Press.
- Bell, J.S. (1964) "On the Einstein-Podolsky-Rosen Paradox", *Physics*, 1:195-200, reprinted in Bell 1987
- Bell, J. (1987) *Speakable and Unspeakable in Quantum Mechanics*, New York: Cambridge University Press.
- Bem, D.J. "Self-perception: An alternative interpretation of cognitive dissonance phenomena" in *Psychological Review* 74: 183-200.
- Benedict, R. (1934/2005) *Patterns of Culture*, New York: Mariner Books.
- Benedict, R. (1934) "Anthropology and the Abnormal," *Journal of General Psychology* 10: 59-82.
- Bernasconi, R. (2002) "Kant as an Unfamiliar Source of Racism" in J. Ward and T. Lott (eds) *Philosophers on Race: Critical Essays*, Oxford: Blackwell.
- Bernasconi, R. (2001) "Who Invented the Concept of Race?" in R. Bernasconi (ed.), *Race*, Oxford: Blackwell, pp. 11-36.
- Bernstein, R., trans. and ed. (2003) *Classical and Romantic German Aesthetics*, Cambridge: Cambridge University Press.
- Bishop, M. and Trout, J.D. (2005) *Epistemology and the Psychology of Human Judgment*, Oxford: Oxford University Press.
- Blackburn, S. (1998) *Ruling Passions*. Oxford: Oxford University Press.
- Block, N. (1980) "What is Functionalism?" In Ned Block (ed.) *Readings in Philosophy of Psychology* vol 1., Cambridge: Harvard University Press, pp. 171-84.
- Bohr, N. (1935) 'Can Quantum-Mechanical description of Physical Reality Be Considered Complete?' *Physical Review* **48**, 696-702.

- Bohm, D., 1951, *Quantum Theory*, New York: Prentice Hall
- Bok, S. (2002) *Common Values*, Columbia, MO: University of Missouri Press.
- Borofsky, R. and Kane, H. (1997) "Cook, Lono, Obeyesekere, and Sahlins" (with replies by Obeyesekere and Sahlins) in *Current Anthropology*, 38:255-282.
- Botterill, G. and Carruthers, P. (1999) *The Philosophy of Psychology*, Cambridge: Cambridge University Press.
- Bowie, A. (1993) *Schelling and Modern European Philosophy: An Introduction*, Routledge.
- Bowie, A. (2003) *Introduction to German Philosophy: From Kant to Habermas*, Cambridge: Polity Press.
- Boxill B. and Hill Jr., T. (2001) 'Kant and Race', in B. Boxill (ed.), *Race and Racism*, Oxford: Oxford University Press, pp. 448-71.
- Brandt, R., ed. (2000) *Rorty and his Critics*, Oxford: Blackwell.
- Brink, D. (1989) *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press.
- Brook, A. (1994) *Kant and the Mind*, Cambridge: Cambridge University Press.
- Brown, D. (1991) *Human Universals*, Philadelphia, PA: Temple University Press.
- Brown, S. (2008) *Moral Virtue and Nature: A defense of ethical naturalism*, London: Continuum.
- Bruner, J.S. and L. Postman (1949) "On the Perception of Incongruity" in *Journal of Personality* 18:206-23.
- Budick, S. (2010) *Kant and Milton*, Cambridge, MA: Harvard University Press.
- Cartwright, N., 1980, 'Do the Laws of Physics state the Facts', *Pacific Philosophical Quarterly*, 61, 75-84.
- Casscells, W., Schoenberger, A., and Grayboys, T. (1978) "Interpretation by Physicians of Clinical Laboratory Results" in *New England Journal of Medicine* 299:999-1001.

- Chisholm, R. (1957) *Perceiving: A Philosophical Study* (Ithaca, New York: Cornell University Press.
- Churchland, P. (1984) *Matter and Consciousness*, Boston: MIT Press.
- Cohen, A. (2009) *Kant and the Human Sciences: Biology, Anthropology, and History* Hampshire, UK: Palgrave MacMillan.
- Cohen, A. (2008) "The Ultimate Kantian Experience: Kant on *Dinner Parties*", *History of Philosophy Quarterly*, 25: 315-336.
- Cooper, David (1990) *Existentialism: A Reconstruction*, Hoboken, NJ: Wiley-Blackwell, 1990.
- Cushing, J. (1994) *Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony*, Chicago: University of Chicago Press.
- Damasio, A. (1994) *Descartes' Error: Emotion, Reason, and the Human Brain*, New York: Penguin.
- Darwin, C. (1987) *Charles Darwin's Notebooks* (ed. P. Barrett, P. Gautrey, S. Herbert, D. Kohn, and S. Smith) Ithaca, NY: Cornell University Press.
- Darwin, C. (1859/1982) *The Origin of Species* (ed. J. W. Burrow) New York: Penguin Classics.
- Darwin, C. (1871/1902) *The Descent of Man*, New York: American Home Library Company. (Available on googlebooks.)
- Darwin, C. (1981) *The Descent of Man, and selection in relation to sex* (ed. J.T. Bonner and R. May), Princeton, NJ: Princeton University Press.
- Dawkins, R. (1976) *The Selfish Gene*, Oxford: Oxford University Press.
- Davidson, D. (1980) *Essays on Actions and Events*, Oxford: Oxford University Press.
- Dean, R. (2006) *The Value of Humanity in Kant's Moral Theory*, Oxford: Oxford University Press.
- de Beauvoir, S. (1987) *Force of Circumstance*, New York: Penguin
- Denis, L. (2007) "From Friendship to Marriage: Revising Kant," *Philosophy and Phenomenological Research* 63:1-28.
- Dennett, D. (2003) *Freedom Evolves*, New York: Penguin.

- Dennett, D. (1995) *Darwin's Dangerous Idea*, New York: Simon and Schuster.
- Dennett, D. (1991) *Consciousness Explained*, New York: Back Bay Books.
- Dennett, D. (1984) *Elbow Room: The varieties of free will worth wanting*, Cambridge, MA: MIT Press.
- Dennett, D. (1978) "Skinner skinned," in *Brainstorms: Philosophical Essays on mind and psychology* Cambridge, MA: MIT Press.
- Derrida, J. (1996) *The Gift of Death*, trans. D. Wills, Chicago: University of Chicago Press.
- deSousa, R. (2010) "Emotion" in *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/emotion/> (accessed January 2012)
- diGiovanni, G. and Harris, H.S., eds. (1985) *Between Kant and Hegel: Texts in the Development of Post-Kantian Idealism*, Indianapolis: Hackett.
- Doris, J. (2002) *Lack of Character: Personality and Moral Behavior*, Cambridge: Cambridge University Press.
- Dreyfus, H. and Rabinow, P. (1982) *Michel Foucault: Beyond Structuralism and Hermeneutics*, Chicago: University of Chicago Press.
- Dutton, D. and Aron, A. (1974) "Some Evidence for Heightened Sexual Attraction under Conditions of High Anxiety" in *Journal of Personality and Social Psychology* 30: 510-517.
- Einstein, A., Podolsky, B., and Rosen, N. (1935) 'Can Quantum-Mechanical description of Physical Reality Be Considered Complete?' *Physical Review* **47**, 777-780.
- Eze, E. (1994) "The Colour of Reason: the Idea of 'Race' in Kant's Anthropology" in K.M.Faull (ed.) *Anthropology and the German Enlightenment*, Bucknell University Press.
- Feldman, R. (2001) "Naturalized Epistemology" in *The Stanford Encyclopedia of Philosophy* (<http://plato.stanford.edu/entries/epistemology-naturalized/>, accessed 3-3-2009.)
- Fichte, J.G. (1993) *Fichte: Early Philosophical Writings* (ed. D. Breazeale) Ithaca: Cornell University Press.

- Fichte, J.G. (1994) *Introductions to the Wissenschaftslehre* (ed. D. Breazeale) Indianapolis: Hackett.
- Fodor, J. (1983) *Modularity of Mind: An Essay on Faculty Psychology*, Cambridge, Mass.: MIT Press.
- Fodor, J. (1968) "The Appeal to Tacit Knowledge in Psychological Explanation," in *Journal of Philosophy* 65: 627-640.
- Foucault, M. (2008) *Introduction to Kant's Anthropology*, trans. R. Nigro and K. Briggs, Los Angeles: Semiotext(e).
- Foucault, M. (2002) *The Archaeology of Knowledge* (trans. A. Sheridan), London: Routledge.
- Foucault, M. (1994) *The Birth of the Clinic*, trans. A.M. Sheridan Smith, New York: Vintage.
- Foucault, M. (1990) *The History of Sexuality, vol. 1*, trans. R. Hurley, New York: Vintage.
- Foucault, M. (1988) *Madness and Civilization*, trans. R. Howard, New York, Vintage.
- Foucault, M. (1984) *The Foucault Reader* (ed. P. Rabinow), New York: Vintage.
- Foucault, M. (1983) "On the Genealogy of Ethics: An overview of a work in progress" in *Michel Foucault: Beyond Structuralism and Hermeneutics* (2d edition) H. Dreyfus and P. Rabinow, Chicago: University of Chicago Press: 229-252.
- Foucault, M. (1982) "The Subject and Power" in *Michel Foucault: Beyond Structuralism and Hermeneutics*, H. Dreyfus and P. Rabinow, Chicago: University of Chicago Press: 208-227
- Foucault, M. (1980) *Power/Knowledge: Selected Interviews and Writings*, ed. C. Gordon, Pantheon Books.
- Foucault, M. (1977) *Discipline and Punish*, trans. A.M. Sheridan Smith, New York: Pantheon.
- Foucault, M. (1973) *The Order of Things*, New York: Vintage.
- Foucault, M. (1969/1982) *The Archeology of Knowledge*, trans. A.M. Sheridan Smith, New York: Vintage.

- Foucault, M. (1961/2006) *A History of Madness* (ed. J. Khalfa, trans. J. Murphy), London: Routledge.
- Frankfurt, H. (1988) *The Importance of What we Care About*, Cambridge: Cambridge University Press.
- Frazier, R.L. (1998) "Intuitionism in Ethics" in *The Routledge Encyclopedia of Philosophy*, ed. E. Craig, London: Routledge.
- Freud, S. (1923/1960) *The Ego and the Id*, (trans. James Strachey) New York: W.W. Norton and Company.
- Freud, S. (1900/1965) *On the Interpretation of Dreams* (trans. James Strachey) New York: Avon Books.
- Freud, S. (1901/1960) *The Psychopathology of Everyday Life*, New York: W.W. Norton and Company.
- Freud, S. (1914) "On Narcissism: An introduction"
- Freud, S. (1920/1963) *Introductory Lectures on Psychoanalysis* (trans. James Strachey) New York: W.W. Norton and Company.
- Freud, S. (1920/1961) *Beyond the Pleasure Principle* (trans. James Strachey) New York: W.W. Norton and Company.
- Freud, S. (1927/1961) *The Future of Illusion* (trans. James Strachey) New York: W.W. Norton and Company.
- Freud, S. (1933/1964) *New Introductory Lectures on Psychoanalysis* (trans. James Strachey) New York: W.W. Norton and Company.
- Friedman, M. (2001) *Dynamics of Reason*, Stanford, CA: CLSI Publications.
- Frierson, P. (2010a) "Two Standpoints and the Problem of Moral Anthropology" in *Kant's Moral Metaphysics* (ed. Benjamin Libscomb and James Krueger, Berlin: deGruyter.
- Frierson, P. (2010b) "Rational Faith: God, Immortality, and Grace" in *Kant: Key Concepts* (ed. W. Dudley and K. Engelhard), Durham, UK: Acumen Publishing.
- Frierson, P. (2010c) "Kantian Moral Pessimism" in *Kant's Anatomy of Evil* (ed. S. Anderson-Gold and P. Muchnik), Cambridge: Cambridge University Press.

- Frierson, P. (2009a) "Kant on Mental Disorder 1: An Overview" in *History of Psychiatry* 20: 267-289.
- Frierson, P. (2009b) "Kant on Mental Disorder 2: Philosophical Implications of Kant's Account" in *History of Psychiatry* 20:290-310.
- Frierson, P. (2008) "Empirical psychology, common sense, and Kant's empirical markers for moral responsibility" in *Studies in the History and Philosophy of Science* 39(4):473-82.
- Frierson, P. (2007a) "Review of Richard Dean, *The Value of Humanity in Kant's Moral Theory*" in *Notre Dame Philosophical Reviews*, at <http://ndpr.nd.edu/review.cfm?id=9364>.
- Frierson, P. (2007b) "Providence and Divine Mercy in Kant's Ethical Cosmopolitanism," in *Faith and Philosophy* 24: 143-63.
- Frierson, P. (2006) "Character and Evil in Kant's Anthropology," *Journal of the History of Philosophy* 44:623-44.
- Frierson, P. (2005) "The moral importance of politeness in Kant's anthropology," *Kantian Review* 9:105-27.
- Geertz, C. (1973) *The Interpretation of Cultures*, New York: Basic Books.
- Geertz, C. (2000) *Available Light: Anthropological Reflections on Philosophical Topics*, Princeton, NJ: Princeton University Press.
- Gettier, E. (1963) "Is Justified True Belief Knowledge?" in *Analysis* 23: 121-123
- Gill, S., et. al. (2006) "Metagenomic Analysis of the Human Distal Gut Microbiome," *Science* 312: 1355-1359.
- Gilligan, C. (1984) *In a Different Voice: Psychological Theory and Women's Development*, Cambridge, MA: Harvard University Press.
- Ginet, C. (1990) *On Action*, Cambridge: Cambridge University Press.
- Ginet, C. (1997) "Freedom, Responsibility, and Agency" in *Journal of Ethics: An International Philosophical Review* 1: 85-98.
- Ginsborg, H. (1997) "Lawfulness without a Law: Kant on the Free Play of Imagination and Understanding" *Philosophical Topics* 25: 37-81.
- Goodenough, W. (1970) *Description and Comparison in Cultural Anthropology*, Chicago: Aldine.

- Gouax, Charles. 1972. "Kant's View on the Nature of Empirical Psychology." *Journal of the History of the Behavioural Sciences* 8: 237-42.
- Grenberg, J. (2005) *Kant and the Ethics of Humility*, Cambridge: Cambridge University Press.
- Gutting, G. (1989) *Michel Foucault's Archeology of Scientific Reason*, Cambridge: Cambridge University Press.
- Gutting, G. (2001) *French Philosophy in the Twentieth Century*, Cambridge: Cambridge University Press.
- Guyer, P. (1987) *Kant and the Claims of Knowledge*, Cambridge: Cambridge University Press.
- Habermas, J. (1990) *Moral Consciousness and Communicative Action*, trans. C. Lenhardt and S.W. Nicholsen, Cambridge, MA: MIT Press.
- Habermas, J. (1993) *Justification and Application*, trans. C.P. Cronin, Cambridge, MA: MIT Press.
- Habermas, J. (2008) *Between Naturalism and Religion*, Cambridge, UK: Polity
- Hamann, J.G. (1800) "Metakritik über den Purismus des Vernunft" in *Mancherley zur Geschichte der metakritischen Invasion*, ed. F.T. Rink, Königsberg, pp. 120-134.
- Hanna, R. (2006) *Kant, Science, and Human Nature*, Oxford: Oxford University Press.
- Harman, G. (2000) *Explaining Value and Other Essays in Moral Philosophy*, Oxford: Clarendon.
- Hatfield, G. (1990) *The Natural and the Normative: Theories of Spatial Perception from Kant to Helmholtz*, Cambridge, MA: MIT Press.
- Hegel, G.W.F. (1991) *The Philosophy of Right* (ed. Allen Wood) Cambridge: Cambridge University Press.
- Hegel, G.W.F. (1979) *The Phenomenology of Spirit* (ed. A.V. Miller) Oxford: Oxford University Press.
- Heidegger, M. (2002) *The Essence of Human Freedom: An introduction to philosophy*, trans. T. Sadler, London: Continuum.

Heidegger, M. (1997) *Kant and the Problem of Metaphysics*, trans. R. Taft, Indianapolis: Indiana University Press.

Heidegger, M. (1985) *Schelling's Treatise on the Essence of Human Freedom* (trans. Joan Staumbaugh), Athens, OH: Ohio University Press.

Heidegger, M. (1982) *The Basic Problems of Phenomenology*, trans. A. Hofstadter, Indianapolis: Indiana University Press.

Heidegger, M. (1977) *Basic Writings*, ed. and trans. D.F.Krell, New York: Harper Collins.

Heidegger, M. (1953) *Sein und Zeit* Tubingen: Max Niemeyer Verlag. (trans. by J. Staumbaugh as *Being and Time*, Albany: SUNY Press, 1996).

Heidegger, M. (2009) *Basic Concepts of Aristotelian Philosophy*, trans. R. Metcalf and M. Tanzer, Indianapolis: Indiana University Press.

Henrich, D. (1955) "The Unity of Subjectivity" trans. G. Zöller in *The Unity of Reason*, ed. R. Velkley, Boston: Harvard University Press.

Herder, J. G. (2002) *Philosophical Writings* (ed. M. Forster) Cambridge: Cambridge University Press.

Herman, B. (1993) *The Practice of Moral Judgment*, Cambridge, MA: Harvard University Press.

Herman, B. (2007) *Moral Literacy*, Cambridge, MA: Harvard University Press.

Horton, J. and S. Mendus eds. (1995) *After MacIntyre: Critical Perspectives in the Work of Alasdair MacIntyre*, South Bend, IN: University of Notre Dame Press.

Husserl, E. (1965) *Phenomenology and the Crisis of Philosophy*, trans. Q. Lauer, New York: Harper and Row.

Husserl, E. (1975) *The Paris Lectures* (trans. P. Koestenbaum and S.J. Bartlett), Dordrecht: Kluwer Academic Publishers.

Husserl, E. (1981) *Shorter Works*, ed. P. McCormick and F.A. Elliston, South Bend, IN: University of Notre Dame Press.

Husserl, E. (1983) *Ideas Pertaining to a Pure Phenomenology and to Phenomenological Philosophy*, trans. F. Kersten, The Hague: Martinus Nijhoff Publishers.

- Husserl, E. (2001) *Logical Investigations*, trans. J.N. Findlay, ed. M. Dummett, London: Routledge.
- Jones, K. (2005) "Moral Epistemology" in *The Oxford Handbook of Contemporary Philosophy*, ed. F. Jackson and M. Smith, Oxford: Oxford University Press.
- Kamtekar, R. (2003) "Situationism and Virtue Ethics on the Content of Our Character," *Ethics* 114: 458-91.
- Kane, R. (1996) *The Significance of Free Will*, New York: Oxford University Press.
- Kane, R. (2002) "Free Will: New Directions for an Ancient Problem" in *Free Will* (Ed. Robert Kane), Malden: Blackwell Publishers, pp. 47-56.
- Kierkegaard, S. (2006) *Fear and Trembling* (ed. C. S. Evans and trans. S. Walsh), Cambridge: Cambridge University Press.
- Kierkegaard, S. (1992) *Concluding Unscientific Postscript* (trans. H. Hong and E. Hong) Princeton, NJ: Princeton University Press.
- Kierkegaard, S. (1989) *Sickness Unto Death*, trans. A. Hannay, London: Penguin.
- Kierkegaard, S. (1985) *Philosophical Fragments* (trans. H. Hong and E. Hong) Princeton, NJ: Princeton University Press.
- Kim, J. (1988) "What is Naturalized Epistemology?" *Philosophical Perspectives* 2 (ed. James E. Tomberlin), Asascadero, CA: Ridgeview Publishing Co: 381-406.
- Kitcher, P. (1993) *Kant's Transcendental Psychology*, Oxford: Oxford University Press.
- Kleingeld, P. (1993) "The Problematic Status of Gender-Neutral Language in the History of Philosophy: The Case of Kant" in *Philosophical Forum* 25:134-150.
- Kleingeld, P. (2007) "Kant's Second Thoughts on Race," *The Philosophical Quarterly* 57:1-20.
- Knobe, J. and Nichols, S., eds. (2008) *Experimental Philosophy*, Oxford: Oxford University Press.
- Korsgaard, C. (2009) *Self-Constitution: Agency, Identity, and Integrity*, Oxford: Oxford University Press.

Korsgaard, C. (2004) *Fellow Creatures: Kantian Ethics and our Duties to Animals*, Tanner Lectures, available at <http://www.people.fas.harvard.edu/~korsgaard/CMK.FellowCreatures.pdf> (accessed 9-15-2010).

Korsgaard, C. (1996a) *Creating the Kingdom of Ends*, Cambridge: Cambridge University Press.

Korsgaard, C. (1996b) *The Sources of Normativity*, Cambridge: Cambridge University Press.

Kramer, P. (2006) *Freud: Inventor of the Modern Mind*, HarperCollins.

Kuehling, S. (2005) *Dobu: Ethics of Exchange on a Massim Island*, Honolulu: University of Hawaii Press.

Kuhn, T. (1996) *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press.

Kuhn, T. (1977) "Objectivity, Value Judgment, and Theory Choice" in *The Essential Tension*, Chicago: University of Chicago Press: 320-39.

Larrimore, M. (1999) "Sublime Waste: Kant on the Destiny of the 'Races'" in C. Wilson (ed.) *Civilization and Oppression*, Calgary: Calgary University Press.

Larrimore, M. (2001) "Substitutes for Wisdom: Kant's Practical Thought and the Tradition of the Temperaments," *Journal of the History of Philosophy* 39: 259-88.

Latour, B. and S. Woolgar, (1986) *Laboratory Life: The Construction of Scientific Facts*, Princeton, NJ: Princeton University Press..

Lessing, G.E. (2005) *Lessing: Philosophical and Theological Writings* (trans. H.B.Nisbet), Cambridge: Cambridge University Press.

Levinas, E. (1969) *Totality and Infinity: an Essay on Exteriority*, trans. A. Lingis, Pittsburgh: Dusquesne University Press.

Levinas, E. (1997) *Basic Philosophical Writings*, ed. A. Peperzak, S. Critchley, and R. Bernasconi, Indianapolis: Indiana University Press.

Levinas, E. (1998a) *Otherwise than Being, or Beyond Essence*, trans. A. Lingis, Pittsburgh: Dusquesne University Press.

- Levinas, E. (1998b) *Entre Nous*, trans. M.B. Smith and B. Harshay, New York: Columbia University Press.
- Libet, B. (1999) "Do we have a free will?" in *The Volitional Brain* (ed. Benjamin Libet, Anthony Freedom, and Keith Sutherland), Thorverton, UK: Imprint Academic.
- Longino, H. (1990) *Science as Social Knowledge: Values and Objectivity in Scientific Knowledge*, Princeton, NJ: Princeton University Press.
- Longuenesse, B. (1997) *Kant and the Capacity to Judge*, Princeton: Princeton University Press.
- Louden, R. (2000) *Kant's Impure Ethics*, Oxford: Oxford University Press.
- MacFarland, J.D (1974) "The Bogus Unity of the Kantian Philosophy," *Proceedings of the Ottawa Congress on Kant in the Anglo-American and Continental Traditions*.eds. Pierre Laberge, Francois Duchesneau, and Bryan E. Momssey. Ottawa: University of Ottawa Press, pp. 180-96.
- MacIntyre, A. (1990) *Three Rival Versions of Moral Inquiry: Encyclopedia, Genealogy, and Tradition*, South Bend, IN: University of Notre Dame Press.
- MacIntyre, A. (1988) *Whose Justice? Which Rationality?* South Bend, IN: University of Notre Dame Press.
- MacIntyre, A. (1984) *After Virtue*, South Bend, IN: University of Notre Dame Press.
- Marx, K. (1963) *Karl Marx: Early Writings* (trans. and ed. T.B. Bottomore) New York: McGraw Hill.
- Marx, K. (1970) *The German Ideology* (ed. and trans. C. Arthur), International Publishers.
- Marx, K. (1977) *Capital, vol. 1* (trans. B. Fowkes) New York: Vintage Books.
- Marx, K. (1994) *Karl Marx: Selected Writings* (ed. L. H. Simon) Indianapolis: Hackett.
- Mayr, E. (1988) *Towards a New Philosophy of Biology: Observations of an Evolutionist*, Cambridge, MA: Harvard University Press.
- McCarty, R. (1993) "Kantian Moral Motivation and the Feeling of Respect" in *Journal of the History of Philosophy* 31: 421-435.

- McCarty, R. (1994) "Motivation and Moral Choice in Kant's Theory of Rational Agency" in *Kant-Studien* 85: 15-31.
- McMullin, E. (1993) "Rationality and Paradigm change in science" in P. Horwich, ed., *World Changes: Thomas Kuhn and the nature of science* Boston: MIT Press, pp. 55-78.
- Mendus, S. (1992) "Kant: An Honest but Narrow-minded Bourgeois?" in H. Williams (ed), *Essays on Kant's Political Philosophy*, Chicago: University of Chicago Press.
- Merleau-Ponty, M. (1945) *Phénoménologie de la Perception*, Paris: Gallimard.
- Moore, G.E. (1913) *Principia Ethica*, Buffalo: Prometheus Books (1968 reprint).
- Moore, G.E. (1959) *Philosophical Papers*, London: Allen and Unwin.
- Munzel, G. Felicitas (1999) *Kant's Conception of Moral Character: The "Critical" Link of Morality, Anthropology, and Reflective Judgment*. Chicago, The University of Chicago Press.
- Murphy, D. and Bishop, M., eds. (2009) *Stich and His Critics*, Hoboken, NJ: Wiley-Blackwell.
- Nagel, T. (1986) *The View from Nowhere*, Oxford: Oxford University Press.
- Nagel, T. (1974) "What is it like to be a Bat?" *Philosophical Review*, 83, 435-456.
- Nietzsche, F. (1997) *Untimely Meditations*, ed. D. Breazeale, Cambridge: Cambridge University Press.
- Nietzsche, F. (1966) *Beyond Good and Evil* (ed. and trans. W. Kaufman) New York: Random House.
- Nietzsche, F. (1967) *On the Genealogy of Morals and Ecce Homo* (ed. and trans W. Kaufman) New York: Random House.
- Nietzsche, F. (1974) *The Gay Science* (ed. and trans W. Kaufman) New York: Random House.
- Nietzsche, F. (1978) *Thus Spake Zarathustra* (ed. and trans W. Kaufman) New York: Penguin.
- Nisbett, R.E. and Wilson, T.D. (1977) "Telling more than we can know: Verbal reports on mental processes" in *Psychological Review* 84:231-259.

- Noddings, N. (1984) *Caring: A Feminist Approach to Ethics and Moral Education*, Berkeley, CA: University of California Press.
- Noe, A. (2009) *Out of Our Heads: Why you are not your brain, and other lessons from the biology of consciousness*, New York: Hill & Wang.
- Novalis (1797) *Miscellaneous Remarks*, abridged trans. in Berstein 2003:203-213.
- Obeyseskere, G. (1993) *The Apotheosis of Captain Cook*, Princeton: Princeton University Press.
- Obeyseskere, G. (2005) *Cannibal Talk: The Man-Eating Myth and Human Sacrifice in the South Seas*, Berkeley, CA: University of California Press.
- Overbye, D. (2007) "Free Will: Now you have it, now you don't" In: *The New York Times*, published January 2, 2007.
- Pereboom, D. (2001) *Living without Free Will*, Cambridge: Cambridge University Press.
- Pickering, A. (1984) *Constructing Quarks: A sociological history of particle physics* Chicago: University of Chicago Press
- Pinker, S. (2007) *The Stuff of Thought: Language as a window into human nature*, New York: Viking.
- Pinker, S. (2002) *The Blank Slate: The modern denial of human nature*, New York: Viking.
- Pinker, S. (1997) *How the Mind Works*, New York: W.W. Norton.
- Pistorius, H.A. (1786) "[Review of] *Erlauterungen uber des Herrn Professor Kant Kritik der reinen Vernunft* von Johann Schultz," *Allgemeine Deutsche Bibliothek* 66:92-123.
- Plantinga, A. (1996) "Darwin, Mind, and Meaning," in *Books and Culture* (May/June, 1996), currently available at <http://www.veritas-ucsb.org/library/plantinga/Dennett.html> (accessed 3-2-2009).
- Plantinga, A. (1993) *Warrant and Proper Function*, Oxford: Oxford University Press.
- Popper, K. (1972) *Objective Knowledge: An Evolutionary Approach*, Oxford: Clarendon Press.

- Putnam, H. (1967) "The 'innateness' hypothesis" in *Synthese* 17:12-22, reprinted as "The Nature of Mental States" in Putnam 1975:429-440.
- Putnam, H. (1975) *Mind, Language, and Reality*. Cambridge: Cambridge University Press.
- Quine, W.V.O. (1990) "Norms and Aims" in *The Pursuit of Truth*, Cambridge, MA: Harvard University Press.
- Quine, W.V.O. (1969) "Natural Kinds," In *Ontological Relativity and Other Essays*, New York: Columbia University Press.
- Quine, W.V.O. (1951) "Two Dogmas of Empiricism" in *Philosophical Review* 60: 20-43
- Rawls, J. (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.
- Reath, A. (1989) "Kant's Theory of Moral Sensibility: Respect for the Moral Law and the Influence of Inclination." *Kant-Studien* 80: 284-302.
- Rescorla, R.A. (1991) "Associative relations in instrumental learning" in *Quarterly Journal of Experimental Psychology* 43b: 1-23.
- Reinhold, K.L. (2006) *Letters on the Kantian Philosophy* (ed. Karl Ameriks, trans. James Hebbeler) Cambridge: Cambridge University Press.
- Rockmore, T. (2002) *Marx after Marxism*, Hoboken, NJ: Wiley-Blackwell.
- Rorty, R. (1990) *Objectivity, Relativism, and Truth (Philosophical Papers, vol. 1)*, Cambridge: Cambridge University Press.
- Rorty, R. (1988) *Contingency, Irony, Solidarity*, Cambridge: Cambridge University Press.
- Rorty, R. (2000) "Universality and Truth" in *Rorty and his Critics*, ed. R. Brandom, Oxford: Blackwell.
- Rumsey, J. (1989) "The Development of Character in Kantian Moral Theory" in *Journal of the History of Philosophy* 27: 247-65.
- Russell, B. (1912/1988) *The Problems in Philosophy*, London: Oxford University Press, 1912, reprinted by Prometheus Books, 1998.
- Said, E. (1979) *Orientalism*, New York: Vintage.
- Sartre, J-P. (1993) *Existentialism and the Human Emotions*, New York: Carol Publishing Group.

- Sartre, J-P. (1989) *No Exit and three other plays*, New York: Vintage Press.
- Sartre, J-P. (1966), "Review of *The Order of Things*," in *L'Arc* 30:87-96, reprinted in *Michel Foucault: Critical Assessments*, ed. B. Smart, London: Routledge, 1994.
- Sartre, J-P. (1956) *Being and Nothingness*, trans. H. Barnes, New York: Simon and Schuster.
- Sartre, J-P. (1947-8/1992), *Notebooks for an Ethics*, trans. D. Pellauer, Chicago: University of Chicago Press.
- Sartre, J-P. (1938) *La Nausée*, Paris: Édition Gallimard.
- Sassen, B. (2000) *Kant's Early Critics*, Cambridge: Cambridge University Press.
- Sayre-McCord, G. (1996) "Coherentist Epistemology and Moral Theory" in *Moral Knowledge? New Readings in Moral Epistemology*, eds. W. Sinnott-Armstrong and M. Timmons, Oxford: Oxford University Press.
- Schelling, F.W.J. (1978) *System of Transcendental Idealism (1800)*, (trans. Peter Heath) Charlottesville: University Press of Virginia.
- Schiller, F. and Goethe, J.W. (1796) *Xenien*, available online at <http://pge.rastko.net/dirs/pge/pge05/100047/e100047.html>, accessed 9-1-10.
- Schiller, F. (1967) *On the Aesthetic Education of Man* (ed. and trans. E. Wilkinson and L.A. Willoughby), Oxford: Clarendon Press.
- Schiller, F. (2006) *Aesthetical and Philosophical Essays*, Middlesex: Echo Library.
- Schlegel, F. (1800) "Ideas" in Bernstein 2003: 261-268.
- Schleiermacher, F. (2006) *On Religion: Speeches to its Cultural Despisers* (ed. and trans. R. Crouter) Cambridge: Cambridge University Press.
- Schmidt, J. (1996) *What is Enlightenment? 18th Century Questions and 20th Century Answers*, Berkeley: University of California Press.
- Schopenhauer, A. (1966) *The World as Will and Representation*, vol. 1 (trans. E.F.J. Payne), Mineola, NY: Dover.
- Schott, R. M., ed. (1997) *Feminist Interpretations of Immanuel Kant*, University Park, PA: Penn State Press.

- Schott, R. M. (1993) *Cognition and Eros: A Critique of the Kantian Paradigm*, University Park, PA: Penn State Press.
- Schwartz, S. and Bilsky, W. (1990) "Toward a Theory of the Universal Content and Structure of Values: Extensions and Cross-Cultural Replications," *Journal of Personality and Social Psychology* 58: 878-91.
- Shell, S.M. (1996) *The Embodiment of Reason*, Chicago: University of Chicago Press.
- Shell, S.M. (2003) "Kant's 'True Economy of Human Nature': Rousseau, Count Verri, and the Problem of Happiness" in *Essays on Kant's Anthropology*, ed. P. Kain and B. Jacobs, Cambridge: Cambridge University Press.
- Sidwick, H. (1901) *The Methods of Ethics*, London: Macmillan & Co.
- Sinnott-Armstrong, W. (1992) "Intuitionism" in *The Encyclopedia of Ethics*, ed. L. Becker, New York: Garland.
- Snow, D. (1996) *Schelling and the end of idealism*, Albany: State University of New York Press.
- Sosa, E. (2009) "A Defense of the Use of Intuitions in Philosophy" in *Stich and His Critics* (ed. D. Murphy and M. Bishop), Hoboken, NJ: Wiley-Blackwell.
- Sperber, D. (1982) "Apparently Irrational Beliefs," in M. Hollis and S. Lukes, eds. *Rationality and Relativism*, Cambridge, MA: MIT Press, pp. 148-80.
- Spiro, M. (1987) *Culture and Human Nature*, Chicago: University of Chicago Press.
- Spiro, M. (1954) "Human Nature in its Psychological Dimensions", *American Anthropologist* 56:19-29.
- Sturgeon, N. (2002) "Ethical Intuitionism and Ethical Naturalism," in *Ethical Intuitionism: Re-evaluation*, ed. P. Stratton-Lake, Oxford: Clarendon.
- Timmons, M. (1999) *Morality without Foundations: A defense of ethical contextualism*, Oxford: Oxford University Press.
- Tolman, E.C. and Hoznik, C.H. (1930). "Introduction and removal of reward, and maze performance in rats." In *University of California Publications in Psychology* 4: 257-275.

- Tye, M. (2007) "Qualia" (revised entry) in *The Stanford Encyclopedia of Philosophy*, available at <http://plato.stanford.edu/entries/qualia/> (accessed June, 2010).
- Valenstein, E. S. (1986) *Great and Desperate Cures*, New York, Basic Books.
- van Cleve, J. (1999) *Problems from Kant*, Oxford: Oxford University Press.
- vanInwagen, P. (1986) *An Essay on Free Will*, Oxford: Oxford University Press.
- Wallace, R. J. (1994) *Responsibility and the Moral Sentiments*, Cambridge, MA: Harvard University Press.
- Wallace, R. J. (2006) *Normativity and the Will: Essays on moral psychology and practical reason*, New York: Oxford University Press.
- Watkins, E. (2005) *Kant and the Metaphysics of Causality*, Cambridge: Cambridge University Press.
- Weinberg, J., Nichols, S., and Stich, S. (2008) "Normativity and Epistemic Intuitions" in *Experimental Philosophy* (ed. J. Knobe and S. Nichols), Oxford: Oxford University Press.
- Westphal, K. (2005) *Kant's Transcendental Proof of Realism*, Cambridge: Cambridge University Press.
- White, A. (1983) *Schelling: An Introduction to the System of Freedom*, New Haven: Yale University Press.
- Wilson, C. (1998) "Savagery and the Supersensible: Kant's Universalism in Historical Context," in *History of European Ideas* 24: 315-330.
- Wilson, E.O. (1978/2004) *Sociobiology*, Belknap Press.
- Wood, A. (1999) *Kant's Ethical Thought*, Cambridge: Cambridge University Press.
- Wood, A. (1998) "Kant on Duties Regarding Nonrational Nature," *Proceedings of the Aristotelian Society Supplement*, Volume 72.
- Wundt, W. (1912) *An Introduction to Psychology* (trans. by Rudolph Pintner) London: George Allen and Unwin.
- Zammito, J. (2002) *Kant, Herder, and the Birth of Anthropology*, Chicago: University of Chicago Press.

Zuckert, R. (2010) *Kant on Beauty and Biology: An Interpretation of the Critique of Judgment*, Cambridge: Cambridge University Press.