Kant's Questions: What is the Human Being?

Patrick R. Frierson

June, 2009

The present version of this book is a draft. Please do not quote without persmission.
Comments are greatly appreciated and may be sent to frierspr@whitman.edu

Chapter 8: Scientific Naturalism

When Kant developed his answer to 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 womb[1] 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).[2] Empirical psychology was based more on introspection than strict scientific methodology and was not yet distinguished from philosophy itself.

Things have changed. Just consider a few of the highlights of our scientific knowledge about human beings[3]:

- 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 has offered evidence that irrationality is widespread and unrecognized even in the most careful and thoughtful human beings
- Situationist psychology provides evidence suggesting that human behavior is determined by context rather than by character

And 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 finally be answered by *science* rather than philosophy. In other words, this scientific progress provides hope for a rigorous *scientific naturalism* applied to the case of 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 own way of dealing with science is not naturalist in this sense, this chapter will also provide Kantian responses to scientific developments of the past 200 years. Given the richness of empirical research on human beings, it is impossible to present a complete naturalist account of human beings in this short chapter, and the task is made even more difficult by the wide range of different naturalist approaches to human nature. But this chapter offers a few highlights of recent empirical research on human beings, some attempts by recent philosophers to use these developments for philosophical accounts of human nature, and some (broadly Kantian) responses to all this. By the end of this brief survey, I hope that simplistic reactions to naturalism will have, at least, 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, so 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) in the world – 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."[4] 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 the specific parts of the frontal and temporal lobes as loci of linguistic activity, a primary projection area in the parietal lobe that controls most moter activity, and the C-fibers in the peripheral nerves of the somatosensory area 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 contruction 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 for faster and more efficient computer, neurons can be networked together to create a an "everything-connected-to-everything" neural network that is capable of building its own connections based on past experience and model human thought.[5] 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 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 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."[6] Such a view represents a strong scientific naturalism, in that there is nothing more to human beings than our (neuro)physiology.[7] 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 1992: 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."[8] 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 imagine other creatures) that feel pain with a different neurobiological architecture (with some other neural element playing the role of C-fibers).[9] 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).[10]

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 brain[11] 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.[12]

Currently, most philosophers of mind reject both substance-dualism and eliminativism in favor of a *functionalist property-dualism* with *token-identity* between mental and physical states.[13] 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, behavioural outputs, and other mental states."[14] 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, without falling into the problem of multiple realizability, since each *type* of mental state can be realized in many different types of brain-state.[15]

To some, Kant's view on the relationship between mental states and brain-states might seem similar to Descartes's substance-dualism. After all, Kant distinguishes between the noumenal thing-in-itself and the phenomenal appearances, and Kant specifically applies this distinction to human beings, who are both transcendentally 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.[16] [17]

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.[18] 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.[19][20] 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 transcendentally.

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 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, Kant argues that "the soul can perceive itself only through the inner sense" (12:35). But inner sense is purely temporal, whereas the physical body is always *spatio*temporal. Thus, the most that physiological explanation could ever do it to explain "the matter that makes possible" mental phenomena (12:35). Mental phenomena as such will always 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. Secondly, 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 really 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 just 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 and 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 of his own, 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 turn out to be falsified given the progress of neuroscience. At present, however, any optimism about neuroscience that would insist that all psychological laws will eventually be translatable into physical laws governing brains 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 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, a variation that deserves more attention. 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 real need for non-physical laws to fully make sense of human (and other living) beings. In both of these 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 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.[21] The normativity problem calls for a different sort of solution than the problems of irreducibility and qualia. Kant ends up opposing 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 even in empirical description 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.

However, even if Kant's overall account of the human being is compatible with (and even strengthened by) contemporary 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).[22] 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."[23]

In contrast to such apparently paradoxical conclusions, a Kantian 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.[24] For Kant, 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 seeing conscious psychological states as correllated 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 never a *temporal* priority. 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.[25] It should come as no surprise that scientists looking for causes of human actions can 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 *precede* those actions and are *undetermined* by physical 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 rather through psycho-pharmaceuticals 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?").[26] 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 such a physical system as a human brain could have come to exist. Without a naturalist account of its origin, the human brain might seem to be a literally miraculous endowment, reflecting some supernatural design (much as computers reflect their human designers). As we saw in chapter three, 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 hereditable 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.

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 diseases such as cardiovascular disease, arthritis, diabetes, and various kinds of cancer.

The project has also fueled a flurry of new research with the aim of creating more specific and targeted treatments for a number of 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. As Daniel Dennett, one of the foremost philosophical popularizers of Darwinism, puts it,

> 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, as Dennett has suggested, 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 some 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 also specifically tied to particular *kind* of naturalist explanations offered by evolutionary theory.

The last 30 years – starting with the publication of E.O. Wilson's *Sociobiology* and Richard Dawkins' *The Selfish Gene* in the mid-70's – has seen the emergence of extremely sophisticated Darwinian accounts of human nature that seek to move beyond the caricatures of evolution as implying that human beings are fundamentally nothing more than clever,

selfish primates. The full 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 can be used to make sense of those 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. Isn't it a central premise of evolution that everyone is out for themselves? In fact not. Even 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."[27] And as our understanding of the processes of evolution grows, cooperative forms of natural selection have emerged as 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 were to move, they had to move themselves. If they were to generate energy, they generated it for themselves. If they were to break down other cells to get organic materials for themselves (i.e., eat), they had to have within themselves the resources to break down those proteins. 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.[28] Evolutionary "fitness" is not merely – nor even primarily – a matter of killing off one's opponents. It can just as easily be a matter of cooperating in particularly effectively ways.[29] In the human case, for example, families (and societies) whose members cooperate tend to outperform those that are constantly at each other's throats, so familial (and social) affection and cooperation has evolved amongst human beings.

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.[30]

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 "though 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. (DM 162)

The point here is that natural selection does not work exclusively on individuals. 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 tend to spend less time in prison that those that are full of betrayers. Thus gangs that can – somehow – develop a tendency not to betray will be more successful than those that do not. In the evolutionary context, genetic[31] 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.

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 emphasis 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 than all genes will code for whatever it is that 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 propogate 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. 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 replicated 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.[32]

Even if evolutionary theory is not committed to a conception of human beings as thoroughly selfish, however, it 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 [its] temporary homes and jump[s] from brain to brain to reproduce."[33] 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.[34] 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 will 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, cf. 173), when a new form of primate emerged – a euprimate or "superprimate," a "hominid with an infected brain, host to millions of cultural symbionts" (173). The "chief enablers" of this revolution "are the symbiont systems known as languages" (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 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 both 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.[35] 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 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. And while there must be *some* genetic basis for these distinctive characteristics of mind, this basis need provide only the most fundamental context for the evolution of memes.

The combination of more sophisticated thinking in general about the way evolutionary theory works (as in the case of altruism) combined with the addition of memes to the general Darwinian framework of evolution has given rise to Darwinian-naturalistic accounts of human freedom and morality. It is worth pointing out, here, that 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,[36] is simply a relic of scientific ignorance.[37] 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 our 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 Kant's key 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 I can discuss 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".[38] 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 to be 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 the 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 provided from different perspectives. 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, of course, 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" that is capable of thinking about itself. And there is no reason that this new entity need serve the interests of the memes that gave rise to it. Moreover, this new entity is precisely a being that *sets its own ends*. And 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."[39]

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 do harm to ourselves or our kind. We have evolved to have 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 own 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. And here, Kant seems to be in trouble. As we noted in chapter two, one of the central claims of Kant's biology[40] 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 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 grass, 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."[41] Today, the assumption of an *intentional* order in nature is not only unnecessary to biology, but would even be a hindrance to biological progress. Moreover, Kant's pessimism about mechanical explanations in biology was linked with his use of the concept of "predispositions," which plays a crucial role in his empirical anthropology. And predispositions are tendencies in organized beings that are simply *taken for granted* in Kant's biological explanations. Kant offers no explanations of how these predispositions *arose*, and he seems to think that the issue of 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 on this planet is central to Darwinian answers to the question, "What is the human being?"[42] At the 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 at least 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. Evolution by natural selection, though not "teleological" in Kant's sense, is a principle for explanation that is distinct from the sorts of mechanical explanation that dominate physics.[43] Even if "natural selection" is in principle explicable in terms of physical forces (the cold kills the less furry animals for physical reasons, and so on), 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. From Darwin to today, biological features selected for are, like Kant's predispositions, typically taken for granted. While molecular biology now gives tools for explaining how genetic (and thus phenotypic) variations arise, such explanations play a relatively minor role in explanations of natural selection.[44]

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."[45]

From the standpoint of the philosophy of biology, Kant's overall approach arguably fits well with contemporary evolutionary biology. Even if Kant was wrong about the specific principles that regulate the practice of biology, he was correct that some heuristic principle of a broadly purposive nature is needed in biology, a principle that is not needed for physical-mechanical explanations of non-living beings. And even if the origins of (human) predispositions are in principle explicable in terms of evolution, Kant was correct that any 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 transcendentally, 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.[46] Because empirical anthropology provides the empirical insights needed for pragmatic anthropology, revisions and additions

to our empirical picture of human beings can also 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 both 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 fails to provide good reasons to deny that transcendental anthropology is both distinct from and more fundamental than 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, then, 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, are new perspectives that 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).

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[47] the relative *priority* of transcendental and empirical anthropology. In particular, 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; 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-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.[48]

Should we be strong scientific realists? There are at least two reasons for skepticism here.[49] 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, for instance, 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 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 in that one could simply take on faith that the world we experience exhausts all that there is.[50] Just as Kant's immediate successors argued for the "neglected alternative,"[51] 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 several 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 even 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 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. To return to an example from chapter two, 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) . . . the 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 (i.e., impossible to estimate). . . 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.[52]

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 certainly correct that nothing about evolution itself provides a reason to believe that beliefs that arise through processes of evolution are *justified* or *true*.[53] Of course, Kant helps us see (contra Plantinga[54]) that evolutionary explanation is wholly *compatible* with an entirely naturalistic account of the origin of human mental faculties. But the fact that evolutionary theory is *compatible* with epistemic justification does not show that it is *sufficient* for it. 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 itself.[55] 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 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, then, 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. To bring up just one 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" here. How would one answer this question? Surely *not* by appealing to our evolutionary history, nor even to the role that

emotions do in fact 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.[56]

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 put to use in a developed pragmatic anthropology. But Kant's transcendental anthropology – or something like it – is necessary in order to explain the conditions of possibility of science itself and to guide the way in 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 have drawn attention to developments in biology that have significant impacts for our understanding of 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. So, 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 may work.[57] But psychology has also made significant progress as a science distinct from biology, most notably through increasingly sophisticated experimental methodologies that provide evidence for various 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 some specific findings in psychology that might seem to raises problems for Kant's anthropology as a whole. This section thus starts with a brief Kantian discussion of some of the methods and models within contemporary psychology and then turns to two ways in which contemporary psychology has fed into naturalist philosophical accounts of human beings in ways that might seem to threaten Kant's epistemology and moral theory.

First, contemporary psychology has made considerably strides towards *re*appropriating 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 expressed in their behavior.[58] 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.[59] 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[60] 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, of course, this is all to be 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. Even though 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, psychologists today are rejecting a "blank-slate" approach to human mental life in favor of an approach that looks a lot like Kant's taxonomy of basic powers.[61] 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."[62] 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.[63]

There are, of course, 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 close to commonsensical or "folk" accounts of mental operations, 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, including specific modules for the perception of color, the perception of shape, the detection of rhythm, and the recognition of other people.[64] 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, of course, 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 be more than satisfied to accept that mental structures that seemed unified to him – such as vision, say, or reason – involved 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."[65] Kant certainly did not draw all of the distinctions drawn by 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 in fact been vindicated by recent psychological research.[66] The fact that many of these processes are unconscious need not pose any intractable problems for Kant's empirical psychology.[67] 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, the lower faculties are not conscious in a reflective sense, and 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 that 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 should aim to explain the full range of human mental processes in terms of the simplest general structures; Specific human 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 (evolutionary) 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, (evolutionary) psychology went 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 contemporary psychology – at least in some forms – is *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 Kant would see as conferring any scientific status. For Kant, mathematics makes physics scientific because it allows the physicist to make *a priori* claims at the foundation of physics. (For example, Kant reasons from 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.) But while mathematical laws of statistics help psychologists process empirical data more effectively, they provide no *a priori* insights into the nature of the mind.[68]

With respect to the second issue – introspection – the divergence between Kant and psychology today might seem more profound. While Kant insisted that empirical anthropology "is provided with a content by inner sense" (7: xxxfrom draft anthro, cf. 25:252, 863-5), contemporary psychologists often disparage "introspection" as an outdated and unscientific approach to studying the mind. Moreover, there is good empirical evidence calling into question 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 for this 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.[69]

Introspection is so unreliable, the authors conclude, 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.[70]

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"[71] 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.[72]

"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 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.[73] 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 the most general terms, these mental states interact with one another. Insofar as psychologists seek to make claims about mental states as such, rather than seeing them as *mere* stimulus-response mechanisms, they must include at least *some* appeal to introspective awareness.[74] [75]

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. Kant 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 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 general approach to psychology, then, is broadly compatible with the methods of contemporary "scientific" psychology. But the success of psychology, like that of neuroscience and evolutionary biology, has also spawned philosophical attempts to appropriate psychology in the service of a thoroughgoing naturalism about human beings. In the case of psychology, this naturalism need be neither materialist nor reductionist, which alleviates some of the problems directed against neuroscientific naturalism in section one. 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 directly defend the "scientific status of . . . psychology directly, without seeking any sort of reduction."[76] 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.[77] [78] 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 -- 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,[79] 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 fact, however, most attempts at naturalism in epistemology today are more modest than Quine's.[80] As J.D. Trout puts it,

> 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.[81]

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.

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 standard forms of irrationality that are pervasive among human reasoners, 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 . . .?[82]

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%).[83] 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."[84] 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 25). Because of the unreliability of basic cognitive processes, we should replace those processes with others that are empirically demonstrated to be more reliable.

With respect to ethics, one prominent use of contemporary research in psychology has been in the service of a "situationist" critique of character based ethical theories.[85] 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."[86] 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 have taken this "situationist psychology" to imply that "Rather 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."[87]

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. The core of Kant's account of cognition is his logic of how people *ought* to think, but most of its empirical detail comes in Kant's attempt to lay out various "prejudices" that determine the ways that people in fact diverge from these ideal ways of thinking. Like the biases and heuristics program, Kant both characterizes the effects of these prejudices and diagnoses their underlying grounds.[88] 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 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 will, of course, require 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,"[89] Kant sees situationist psychology as an empirical confirmation of humans' radical evil. 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,[90] 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 might focus on the 10% of "high hurry" 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" (BT 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%" (BT 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.

Now 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 are really suggesting revising epistemic standards away from reasoning based on what seems to be good evidence towards reasoning that brings about certain good effects, whether aletheic (truth-conducing) or otherwise. And Bishop and Trout in fact suggest just such a revision: "The primary aim of epistemology, from our perspective, is to provide *useful,* general advice about reasoning" (BT 94, emphasis added), where advice is "useful" towards the end of "human well-being," including such things as "avoid[ing] pain and misery" (BT 94); "health, deep social attachments, personal security, the pursuit of significant projects" (BT 99), and even "discovering the truth about the basic physical or social structure of the world" (BT 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.[91] 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.[92] But it should at least be clear that the application of psychological research on the aletheic and practical benefits of certain reasoning strategies is not *sufficient* to justify those strategies epistemically.[93]

But what should we do if we find that reasoning strategies that are most effective for getting accurate and useful results 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 will 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 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 even offered some conjectures 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 with 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. (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.[94] The fact that sciences can explain human behavior causally need not imply (as Kant showed in his Third Antinomy, see chapter two) 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 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.[95]

(2) Indeterminism.[96] In fact, contemporary natural sciences, unlike the Newtonian science of Kant's day, do *not* think that everything in the universe is caused; quantum mechanics postulated genuine indeterminism in nature, and the complexity of the human brain can give rise to contexts within which this quantum indeterminism can make a significant difference 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: pp. 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 some good reason to identify with the quantum fluctuations in my brain rather than the deterministic processes shaped by my genetic and environmental background, the "ends and purposes" that arise from those quantum fluctuations will be no more my own that those that arise from deterministic influences. Because our self-image is largely shaped by features of ourselves that are stable or at least consistent with our past personality, changes that arise from quantum fluctuation may even come to seem *less* my own that 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 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 fluctation 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 the 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.[97] 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.[98] 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 of moral responsibility, and then see whether this is compatible with science.[99]

(4) Perspectivism. The previous three approaches to freedom all fit well with strong scientific realism.[100] 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" (45). For perspectivists of this stripe,[101] 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. (55)[102]

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 morality.[103]

Instead of a science-first perspectivism, then, one might adopt a *neutral perspectivism*, as has become 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."[104] 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 priveleged 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 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,[105] 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.[106] 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.[107]

V. Conclusion.

In the end, where does modern science leave us with respect to the fundamental question, "What is the human being?" 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 the freedom needed to make sense of normativity from-within. 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 ideals for ourselves from-within.

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 11, we will look at one of the twentieth century's defining "from-within" approaches to human beings: existentialism. And in chapter 12, we will 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. Whereas Kant saw these differences as primarily a small sub-field within empirical anthropology, recent thinkers have sought to extend 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.

Summary

Further Reading

Bishop and Trout.
Dawkins
Dennett xxx and xxx
Doris.
Fodor 1983. Classic statement of modular theory of mind.
Kornblith, H., ed. (1997) *Naturalizing Epistemology*, Cambridge, MA: The MIT Press. xxx
Pinker 1997, xxx, and xxx.
Quine, W. V. O. (1969) "Epistemology naturalized." In *Ontological relativity and other essays,* New York: Columbia University Press, pp. 1-21.
Tversky, A., and D. Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* 211 (4481): 453-8.
———. 1982. Judgments of and by representativeness. In *Judgment under uncertainty: heuristics and biases*, edited by D. Kahneman, P. Slovic, and A. Tversky, 84-98. New York: Cambridge University Press.
Situationism textbook xxx.

[1] Perhaps even more dramatically, others argued against this view with purported observations in microscopes of very tiny human forms seen within male spermatozoa.
[2] And Kant insisted, in his Anthropology, that in determining whether someone is legally insane, "the court cannot refer him to the medical faculty but must refer him to the philosophical faculty" (7:213) because philosophy, not medicine or biology, was the primary context for studying what we now call "psychology."
[3] In putting together this list, I have ignored developments in quasi- and pseudo-scientific fields such as psychoanalysis, xxx, and xxx. For the sake of simplicity, I focus in this chapter on human sciences widely recognized as scientific. A good standard here is the membership of the American Academy of Sciences, which includes many psychologists (including clinical psychologists) but, to the best of my knowledge, no psychoanalysts.xxx check this xxx.
[4] Valenstein, E. S. (1986) *Great and Desparate Cures,* New York, Basic Books, p. 90. For further discussion, see Damasio 1994.
[5] For a helpful discussion of neural networks in relation to computer networks, see Pinker 1997: 99-131.
[6] Paul Churchland (1984) *Matter and Consciousness*, Boston: MIT Press, p. 29.
[7] 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 (see xxx), or even beyond the body in the world in which we live (see Noe and Dennett).

[8] Nagel xxx: xxx. For recent discussions of qualia, see xxx and xxx.

[9] **Putnam, Hilary. (1967). "xxx" reprinted as "The Nature of Mental States" in Putnam (1975) Mind, Language, and Reality. Cambridge: Cambridge University Press, pp. xxx.**

[10] These examples are different 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 predicate certain properties of mental states that do not seem translatable, even in principle, into anything that could be said of a brain-state.

[11] He even famously located the "pineal gland" in the brain as the locus of higher cognitions in human beings.

[12] (and, ultimately, the soul was indestructible and therefore would survive even after being separated from the body)xxxexpand fn.

[13] See Botterill and Carruthers, especially chapter one.

[14] Ned Block (1980) "What is Functionalism?" In Ned Block (ed.) *Readings in Philosophy of Psychology* vol 1., Cambridge: Harvard University Press, pp. 171-84. Block (1980) also collects several other influential papers on functionalism in the philosophy of mind (see pp. 185-306). For an influential early version of functionalism, see Putnam, Hilary. (1967), "Psychological Predicates," in W.H. Capitan and D.D. Merrill, eds., *Art, Mind, and Religion*, Pittsburgh: University of Pittsburgh Press; reprinted as "The Nature of Mental States" Putnam (1975) *Mind, Language, and Reality*, Cambridge: Cambridge University Press.

[15] Re: the type-token distinction, see Davidson 1970.

[16] 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 actions, we can identify a free noumenal ground for those actions. 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.)

[17] For detailed discussions of Kant's philosophy of mind, see Ameriks xxx, Aquila xxx, and Broad xxx.

[18] Whether this requires a metaphysical distinction between two different "things" – a mind-in-itself and an empirical-mind – depends upon one's interpretation of Kant's transcendental idealism. At least, Kant holds that there are different standpoints that one can adopt when considering the mind, and the distinction between transcendental and empirical anthropology in chapters 2 and 3 elaborated these different standpoints in detail. On some readings of Kant's transcendental idealism, the realm of things-in-themselves is literally a different world from the world of appearances, in which case it would be natural to see a distinction between two different minds that somehow relate to one another. However, even most proponents of a "two-world" reading of Kant's transcendental idealism do not think that Kant needs to be committed to a distinction between two different selves, or minds, arguing

instead that Kant can distinguish between two different aspects of a single mind (see Ameriks recent literature essay. Xxx)

[19] For the sake of simplicity, I here conflate two importantly different distinctions: between the empirical and transcendental perspectives and between the empirical self and the self-in-itself.

[20] For a discussion of different approaches to the metaphysics of the transcendental-empirical distinction, see Ameriks "Recent Theoretical Work"xxx. It is noteworthy here that even as Ameriks raises serious problems for non-metaphysical readings of Kant's transcendental idealism (according to which there are simply two different standpoints with no metaphysical commitments xxx), he tacitly endorses a "two-aspect" rather than "two-substance" reading of human agents (xxx).

[21] With respect to normativity, the problem with eliminativism is not 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.

[22] Libet, Benjamin (1999) "Do we have a free will?" in: *The Volitional Brain* (Ed. Benjamin Libet, Anthony Freedom, and Keith Sutherland), Thorverton, UK: Imprint Academic.

[23] OVERBYE, DENNIS (2007), "Free Will: Now you have it, now you don't" In: New York Times, Published January 2, 2007, online at http://query.nytimes.com/gst/fullpage.html?res=9E0CE7D61630F931A35752C0A9619C8B63 (accessed 6-1-09),

[24] Daniel Dennett 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 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.

[25] What would it even mean for the choice to come temporally first? What would the choice itself "look like"? Do we have any reason to think that our observation of a particular mental state – a "decision-event" – must precede our observation of the physical state that is its correlate? Why? Why think that the observed decision-event is identical to (or the observable correlate of) what we decide from-within, rather than thinking that the brain-state is the nearest observable correlate of the "real" – from-within – decision.

[26] As we will see in section 3 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.

[27] (chap. III, ¶3, p. 72 in princeton edition)

[28] One can even trace maternal descent through these mitochondrial DNA because, unlike humans' nucleic DNA, mitochondrial DNA comes entirely from one's mother.

[29] For a more detailed discussion of this, see Dennett Darwin's dangerous idea xxx, xxx

[30] 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.

[31] For the sake of simplicity, I here focus on genetic mutations. Once we introduce the concept of a meme, it should be clear that any mutation that enhanced cooperation within a society, whether that mutation is genetic, memetic, or something else, will enhance the fitness of that society relative to others.

[32] The language of groups "competing" or genes "striving" is obviously anthropomorphic. I mean only that genes are subject to forces of natural selection.

[33] Dennett 2003:175, cf. Dennett xxxx (DDI): xxx and Dawkins 1976: xxx.

[34] 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. □)

[35] Of course, 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.

[36] Of course, even these Darwinian approaches accomodate the phenomenon of morality, the fact that people *take* certain things to be right or wrong, but they deny that these moral norms have real legitimacy. Moral norms are *merely* successful memes.

[37] See, for example, Will Provine xxx, xxx

[38] See B131ff. and the discussion of the transcendental unity of apperception in chapter one.

[39] See Dennett xxx, the title of which is Elbow Room: Varieties of Free Will Worth Wanting.

[40] For the sake of simplicity, I am conflating Kant's philosophy of biology with his biology proper. xxx

[41] Ernst Mayr (1988) Towards a New Philosophy of Biology: Observations of an Evolutionist, Cambridge: Harvard University Press, p. 58.

[42] 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" (See xxx, xxx secondary sources on species debate.). 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 beings are. However, given that Darwin himself calls the "immense" difference in mental power "between the highest ape and the lowest savage" (DM, 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 from a Darwinian point of view.

[43] 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 discussions of teleology in biology, see xxx.) Even if teleology in the strict sense is not required, though, Darwinism confirms the fruitfulness of employing non-mechanistic principles of judgment in biology.

[44] When it comes to memetic explanations offered by philosophical Darwinians, the connection with physical explanation is even more tenuous.

[45] See chapter three, pp. xxx, and chapter five, pp xxx.

[46] It can even help us avoid some of Kant's misuse of teleology in his system, such as at the beginning of the Groundwork, by showing the way in which even moral reason has natural purposiveness. One might think that Kant's empirical claim that humans have a predisposition to personality comes into conflict with contemporary evolutionary theory, but once we enrich that account with memetics, we can accommodate even a predisposition (such as personality) that does not serve one's biological fitness.

[47] both the nature of the transcendental (from-within) perspective and about

[48] For an example of this from Kant's own work, see A173-4/B215-6 (quoted in chapter 10, p. xxx).

[49] In the next chapter, I will emphasize a third reason in the context of recent historicist accounts of science, and in chapter 11, I will look at some existentialist critiques of science.

[50] Put another way, strong scientific realism is a form of what Kant calls "transcendental realism," the view that the world-as-we-experience-it corresponds to, or is identical with, the world-as-it –is-in-itself.

[51] See chapter 6, pp. xxx/.

[52] Alvin Plantinga, "Darwin, Mind, and Meaning," originally published in Books and Culture (May/June, 1996), currently available at http://www.veritas-ucsb.org/library/plantinga/Dennett.html (accessed 3-2-2009).

[53] Moreover, evolutionary theory has at its disposal an account of human cognition that may provide justification for scientific realism. Karl Popper, in Objective Knowledge: An Evolutionary Approach (Oxford: Clarendon Press, 1972), p. 261ff, argues that theories that survive criticism are not only better replicators but also more likely to be *true*, since criticism, at least of the right sorts, aims to sort out views that are not true. And Quine has argued that "creatures inveterately wrong in their inductions have a . . . tendency to die before reproducing" ("Natural Kinds," In Ontological Relativity and Other Essays (New York: Columbia University Press, 1969), p. 126). For criticism of Popper and Quine, see Alvin Plantinga (1993) Warrant and Proper Function, Oxford: Oxford University Press.) However, although evolutionary theory might be able to provide for *some* connection between memetic fitness and truth, it better supports a modest realism according to which our beliefs are true *enough* but reflect the world as it must appear to us in order for us to thrive, not the world as it is "in itself."

[54] Plantinga goes so far as to say, "Modern science was conceived, and born, and flourished in the matrix of Christian theism. Only liberal doses of self-deception and double-think, I believe, will permit it to flourish in the context of Darwinian naturalism." For Kant, however, insofar as Darwinian naturalistic explanations are complemented by a transcendental anthropology of human cognition, science can flourish perfectly well without Christian theism.

[55] I'm not here denying the possibility of a sort of Quinean holism or reflective equilibrium model. 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.

[56] Moreover, Kant actually 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 the problem of radical evil, Kant insists that one must not give up the ideal of moral perfection as an ideal, but also that one can justify moral hope on the grounds of mere progress towards this ideal rather than perfect conformity with it.

[57] For details, consult any recent psychology textbook, or the texts mentioned at the end of this chapter (Pinker).

[58] See the classic study by E.C. Tolman and C.H. Hoznik (1930). "Introduction and removal of reward, and maze performance in rats." University of California Publications in Psychology 4: 257-275. See too studies supporting a cognitive account of learning by R.A. Rescorla (1991) "Associative relations in instrumental learning" (Quarterly Journal of Experimental Psychology 43b: 1-23).

[59] Even the most ardent behaviorists (such as Skinner himself) implicitly made use of descriptions of stimuli and responses that appealed to internal psychological states. See Pinker 1997: 63 and references to Chomsky, Fodor, and Dennett.

[60] Pinker's e.g., see Pinker 1997: 62.

[61] For the sake of simplicity, I am conflating a blank-slate approach with an anti-modular approach here. 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's discussion of these approaches in xxx pp. Xxx-xxx.

[62] Botterill and Carruthers 1999:50. For detailed discussion of modularity, see Fodor 1983 Modularity of Mind: An Essay on Faculty Psychology. Cambridge, Mass.: MIT Press. And for a wholesale assault on the model of the midn as a blank slate, see Pinker xxx.

[63] See e.g. Botterill and Carruthers, p. 96.

[64] Fodor 1983: 47-8.

[65] Botterill and Carruthers, p. 73.

[66] Re: tri-partite, see especially recent work on xxxchemical basisxx of pleasure, which is often but not always linked to desire xxx. Re: higher vs. lower, see xxx. 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, xxx (love on bridge example, insomnia example, xxx).

[67] 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).

[68] Arguably, of course, even modern physics no longer has an a priori foundation. See my discussion of historicism in the philosophy of science in chapter ten.

[69] Nisbett and Wilson 1977: 243-4.

[70] Literally 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 xxxx.)
[71] Nisbett and Wilson 1977: 231.
[72] NW 1977: 238.
[73] 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.
[74] 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. Thus introspection – if genuinely elicited – might be more effective than these experiments suggest.
[75] Functionalist approaches to mental states might seem to provide a third way here, neither behaviorist nor depending upon introspective awareness. However, even functionalism ultimately either reduces to a complex behaviorism or depends upon some level of introspection. Insofar as mental states are treated exclusively as functions that take either stimuli or other functions as their inputs and yield either responses or other functions as their outputs, functionalism is a complex sort of behaviorism. But one can also take mental-state-functions to be not merely functions but also possible objects of experience. (Think here about how the heart, for instance, can be treated functionally – as that into which de-oxygenated blood flows and out of which oxygenated blood flows – but also as an object of experience in its own right.) If one treats mental states as also objects of experience in their own right, the relevant "experience" is fundamentally introspective. If there is more to the claim that "X causes me to desire A" than merely that "X reorganizes my mental architecture in such a way that given Y, I will behave in such-and-such a way," then "desire" gets at least some of its meaning from introspection.
[76] Botterill and Carruthers, pp. 1, 186-7.
[77] An excellent overview is provided in the entry by Richard Feldman on Naturalized Epistemology for the Stanford Encyclopedia of Philosophy (http://plato.stanford.edu/entries/epistemology-naturalized/, accessed 3-3-2009.)
[78] The two main "naturalist" approaches to ethics are evolutionary and Aristotelian. The former was discussed in the previous section. The latter fits oddly in the present chapter, since Aristotelian forms of naturalism are notoriously difficult to reconcile with contemporary natural sciences (especially

Darwinism). For one attempt to integrate Aristotelian virtue ethics with contemporary scientific research, see xxx NDPR review.

[79] For an important critical assessment of this view, see Jaegwon Kim 1988 (e.g. p. 390). Jaegwon Kim 1988 "What is Naturalized Epistemology?" Philosophical Perspectives 2 edited by James E. Tomberlin, Asascadero, CA: Ridgeview Publishing Co: 381-406.

[80] Even Quine's own naturalism has grown more modest. Cf. Quine 1990. "Norms and Aims" in The Pursuit of Truth, Cambridge: Harvard University Press.

[81] 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.

[82] xxxxRef. Cf. Bishop and Trout 122.

[83] 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%.

[84] B and T, especially pp. 104-118.

[85] 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.

[86] Doris Lack 2, cf. Doris, 5-6, 15-22 and Harman Explaining Virtue 168, 178. For an excellent defense of ancient virtue ethics against these sorts of critiques, see Rachana Kamtekar, "Situationism and Virtue Ethics on the Content of Our Character," Ethics 114 (2003): 458-91.

[87] Lack, 146.

[88] For details, see chapter three and my Kant's Empirical Psychology, chapters xxx and xxx.

[89] Lack, 146.

[90] See chapters 2 and 4, and Frierson xxx.

[91] 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.

[92] Xxx Ask Carole and/or Rebecca if there are proponents of a similar position in epistemology.

[93] 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 (BT 156). (Unfortunately, they mistakenly think that defining human well-being is largely an empirical matter (BT 99, 156).

[94] Actually, it involves at least one more non sequitur, since even if human beings are not free, one needs to show that the lack of freedom absolves them of moral responsibility.

[95] 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.

[96] 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 (On Action. Cambridge: Cambridge University Press, 1990 and "Freedom, Responsibility, and Agency." Journal of Ethics: An International Philosophical Review 1.1 (1997): 85-98. For a defense of causal indeterminism, see Kane (The Significance of Free Will. New York: Oxford University Press, 1996) "Free Will: New

Directions for an Ancient Problem." Free Will. Ed. Robert Kane. Malden: Blackwell Publishers, 2002, 47-56.

[97] Xxx Frankfurt, Dennett, Wallace.

[98] The argument for this is discussed in chapter two. Briefly, Kant argues from the fact that the moral law is a categorical imperative to the fact that following the moral law requires acting on xxxxx flesh out problema I from kpv.xxx

[99] For some recent attempts to do just this, see Harry Frankfurt 1998 and R. Jay Wallace xxx and xxx. Even many contemporary Kantians have backed off a strong commitment to transcendental freedom as a condition of possibility of moral responsibility. See Korsgaard 1996.

[100] 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 here emphasize as strongly realist all of those views of freedom that are compatible with strong scientific realism.

[101] 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.

[102] Dennett's claim that "the creation of . . . new standpoints" is "the . . . product of . . . [a biological] revolution" (Dennett 2003: 179) also fits here.

[103] 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*.

[104] Korsgaard, *Sources* 1996: 96.

[105] Actually, he offers at least one other "reason," though it is not exactly a reason to believe that we are free. Aesthetic experience, especially our experience of the sublime, allows us to feel our freedom.

[106] 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.[106] (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.

[107] This freedom-first perspective does not mean, of course, that scientists can simply write off human behavior as inexplicable, nor that if science looks hard enough, it will find some free agent sitting in the control room of our brains. For Kant, the claim that human beings are free has no empirical meaning at all. Instead, it means that scientific explanations are limited to explaining the world in terms of a causal order that fits with our modes of cognition, but that there are standpoints we take on our thoughts and actions – especially "from-within" those thoughts and actions – that reveal aspects of what it is to be human that science cannot explain. Most basically, the postulation of freedom just

means that human beings are responsible for our thoughts and actions, where this responsibility is not an empirical but an epistemic/moral fact. For more on freedom-first perspectivism, see Frierson xxx (Two Standpoint paper).