

### 3 Kant on Human Evil and Human History

In the last chapter, we saw Kant's detailed empirical anthropology. While this anthropology does not rise to the level of a "science" in Kant's strict sense, it is a highly systematic account of universal human characteristics. This chapter looks at two further and related aspects of Kant's empirical account of human beings that flesh out Kant's empirical anthropology and complete unfinished business left by the *Critique of Judgment* regarding what we may hope for humanity (see 11: 429). First, we look at Kant's account of human evil. For Kant, human beings are "radically" evil "by nature." Despite this apparently glum assessment, however, Kant endorses a realistic hope for human goodness. Second, we look at one component of this hope, Kant's philosophy of human history, beginning with the emergence of human beings as a new kind of animal with a rational nature and progressing toward a future of perpetual peace amongst nations and increasingly cosmopolitan political, ethical, and social lives.

#### Radical Evil in Human Nature

##### *"The Human Being is Evil by Nature"*

Kant discusses human evil in his *Anthropology* (7: 324f.) and in various lectures and notes on ethics, anthropology, and religion, but his most sustained discussion of it takes place in *Religion within the Boundaries of Mere Reason*, a work in which Kant aims "to make apparent the relation of religion to a human nature partly laden with good dispositions and partly with evil ones" (6: 11). Kant's argument for human evil is complicated because of apparently contradictory claims. At times, Kant seems to rule out knowing anything about one's moral status at all, saying that "we can never, even by the strictest examination, completely plumb the depths of the secret incentives of our actions" (4: 407; see also 6: 36–37, 63; 8: 270). But Kant does argue for human evil, and when he does so, he claims both that evil "can only be proved [by] anthropological research" and "experiential demonstrations" (6: 25, 35) and that "the

*Kant on Human Evil and Human History* 73

judgment that an agent is an evil human being cannot reliably be based on experience” (6: 20). Insofar as he does appeal to experience, Kant sometimes seems to argue directly from “the multitude of woeful examples that the experience of human deeds parades before us” (6: 32–33), but elsewhere insists that his claim that “the whole species” is evil can be justified only “if it transpires from anthropological research that the grounds that justify us in attributing ... [evil] to human beings ... are of such a nature that there is no cause for exempting anyone from it” (6: 25). From a quick look at these passages, it becomes unclear whether there can even be an argument for human evil, and among Kantians who find such an argument, there is a vibrant debates between those who think that this argument is a priori (e.g. Allison 1990 and 2001) and those who think that it is empirical (see Wood 1999: 287, Frierson 2003).

Fortunately, things are not as hopeless as they seem, and Kant’s various statements can be put together into a complicated but plausible anthropological defense of human evil. The key to putting together Kant’s argument comes at the beginning of *Religion*:

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

(6: 20)

Kant’s argument for evil involves both an empirical component (the experience of “evil actions”) and an a priori component that justifies the inference from these to the “evil maxim” that underlies them. The rest of this section unpacks this argument.

The quotation above implies that one can infer maxims from actions. While this might seem to contradict the claim above about the impossibility of self-knowledge, Kant is actually remarkably consistent. Whenever Kant emphasizes the inscrutability of humans’ motives, he emphasizes only that we can never know that our maxims are *good*. With moral evil, the case is different. While there are no actions that cannot be done from bad motives, there are some actions that cannot be done from good motives. Kant’s reference, in the above quotation, to “actions that are evil” and his specification of these as “contrary to law,” is important. Generally, maxims rather than actions are good or evil. But there are “actions ... contrary to duty” (4: 397), and in his *Metaphysics of Morals*, Kant articulates a political theory based on the intrinsic wrongness of actions that cannot “coexist with everyone’s freedom in accordance with a universal law” (6: 231). Because these actions are wrong regardless of

74 *Kant on the Human Being*

their ends, one can legitimately infer bad underlying maxims from the performance of such actions. Moreover, because moral inscrutability comes partly from humans' tendency to self-flattery, it is implausible that one would pretend to a motive less noble than one's actual motive, so when one finds an evil motive, one can reasonably trust that there is no underlying righteous motive. Motivational inscrutability is asymmetrical: one can never know that a person – including oneself – is morally good, but one can know that people are evil.

Even if Kant's claims about inscrutability do not preclude knowledge of human evil, though, how can Kant make inferences from experience to the existence of human evil given that "the judgment that an agent is an evil human being cannot reliably be based on experience" (6: 20)? Neither experience nor a priori arguments alone are sufficient for Kant's proof of evil. Experience of actions contrary to duty would not be sufficient for ascribing an evil will to human beings without an argument that links those actions to evil maxims. But given evil actions, one knows that if those actions are grounded in freely chosen maxims, then the maxims are evil. So to connect evil actions and evil maxims, all that is needed is an argument that human beings are free agents who choose in accordance with maxims that can ground evil actions such as those found in experience. Kant's transcendental anthropology has already shown that human actions are phenomenal expressions of noumenal, free choices. In *Religion*, Kant adds an account of the specific structure of the fundamental maxim that grounds evil actions.

In particular, *Religion* makes two important additions to the account of free choice found elsewhere in his Critical philosophy. First, he argues that human choices must be grounded in a basic maxim that is either fundamentally good or fundamentally evil; no middle ground is possible.

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

(6: 24–25)

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

Second, Kant connects his transcendental account of humans' free finitude with his empirical account of human predispositions. As we saw in the last chapter, Kant's conception of a "predisposition" has wide application, covering all basic human powers and the instincts and propensities

*Kant on Human Evil and Human History* 75

that direct the faculty of desire. In *Religion*, Kant employs this notion of a predisposition to discuss a fundamental “predisposition to the good” that consists of three distinct “elements of the determination of the human being” – animality, humanity, and personality (6: 26–27). The predisposition to animality includes instincts for self-preservation, sex, and “community with other humans” (6: 26). The predisposition to humanity includes innate tendencies to compare ourselves with others and “inclination[s] to gain worth in the opinion of others” (6: 27). Finally, the predisposition to personality is “susceptibility to respect for the moral law as of itself a sufficient incentive in the power of choice” (6: 27).

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

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

(6: 36, see also 6: 32)

In this important passage, Kant lays out the essence of his account of human evil. Importantly, the account can be read both in terms of transcendental freedom and in terms of empirical anthropology. The transcendental reading is crucial since in the absence of a transcendental perspective on the subordination of moral to non-moral incentives, no empirical claim can imply anything about human evil: “In freedom alone is evil” (18: 212). From the perspective of freedom, when one looks at one’s action from-within, what Kant claims here is that in all choices, we have concern both for morality and for well-being (animal and social inclinations), but that ultimately, we subordinate one concern to the other. Our free (noumenal) nature is constituted by whether we unconditionally prioritize the moral law to non-moral concerns or whether we allow non-moral concerns of sufficient weight to trump the moral law. This aspect

## 76 *Kant on the Human Being*

of Kant's account depends crucially upon the account of morality from Kant's transcendental anthropology, within which Kant shows both that human beings are transcendently free and morally obligated. Here, Kant uses these claims to argue that because morality requires *unconditional* obedience from a transcendently free will, *any* subordination of moral to non-moral concerns is wholly evil.

But Kant's argument for human evil is not merely directed toward helping readers recognize evil from-within. He also makes an empirical-anthropological point – that human beings are evil *by nature*. The quotation above thus helps complete Kant's empirical anthropology. Human beings have various predispositions that can be classified in terms of animal instincts, social inclinations for recognition, and moral interests. But a complete empirical account of human beings must discern how these needs interact in cases when more than one is active. And Kant sees empirical evidence suggesting that the empirical character of human volition is structured such that moral grounds are inactive when they conflict with sufficiently strong non-moral grounds. Kant finds such evidence in the “multitude of woeful examples” of human misdeeds, which shows not only that humans have predispositions that make evil possible but also that we have a volitional structure in which the moral predisposition is made inactive by sufficiently strong sensuous incentives. Given our transcendental freedom (established by Kant's transcendental anthropology), human beings are thus evil. Transcendentally speaking, there is no necessity for human beings to have this volitional structure; it is contingent upon transcendently free choice. But empirically speaking, when one seeks to discern human nature based on empirical evidence, there is good reason to think that human volition subordinates *pure* higher volition to *impure* higher volition. And given that Kant's transcendental anthropology shows this empirical character to be grounded in free choice, there is reason to describe this subordination as “evil.”

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

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

5. Thus, human behavior is characterized by a prioritization of non-moral predispositions over the moral predisposition.
6. Humans' empirical behavior and character express their transcendently free choices.
7. Thus, human beings are morally evil.

The first three premises are empirical generalizations, of different levels of complexity. The first is a straightforward generalization of observations about human beings. The second and third generalize an empirically grounded anthropological explanatory model. These premises are developed in much greater empirical detail, as we showed in Chapter 2. The fourth premise is a moral premise, a part of Kant's a priori, transcendental anthropology of volition. The evidence for this claim is thus a priori. If this a priori premise is taken as stipulative, the preliminary conclusion at (5) could be taken as an empirical–anthropological conclusion. That is, if prioritizing the moral predisposition involves consistency (by definition), it is clear from premises (1)–(3) that human beings act according to a complex structure of predispositions within which the “moral” predisposition is subordinated to others. In that sense, (5) is an empirical fact. But premise (6) is essentially transcendental; there is no empirical evidence for humans' status as free grounds of their empirical characters. Given this premise, however, the prioritization of non-moral predispositions over the moral predisposition that was shown to be a part of human nature is also revealed as an expression of moral evil. The conclusion which is both transcendental and empirical, is that human beings are evil by nature, that is, that moral evil can be ascribed to every member of the human species.

### *The nature of radical evil*

Having shown that human beings are evil, Kant elaborates on the nature of evil. Most importantly, Kant emphasizes that human evil is “radical” in that “it corrupts ... the subjective supreme ground of all maxims” (6: 37). The “maxim” by which humans subordinate moral to non-moral incentives is their most fundamental maxim. In general, humans act in accordance with various principles (maxims) of action, which can be ordered hierarchically. To take one of Kant's own examples, one might act on the maxim “when I believe myself in need of money I shall borrow money and promise to repay it, even though I know that this will never happen” (4: 422), but this maxim is merely a particular application of more general maxims such as “I will trust my own assessments of my needs” and “whenever I can make use of others to satisfy my needs, I will do so,” and this latter maxim is a more specific application of an even more general maxim that

## 78 *Kant on the Human Being*

Kant explains in terms of the relative subordination of inclinations and morality, something like “I will obey the moral law only insofar as doing so is compatible with satisfying other desires, and I will seek to satisfy some non-moral desires.” This last maxim is the fundamental guiding maxim of an evil human being, and all other maxims are merely applications to particular cases where inclinations and/or the moral law are in play. Because this corrupt maxim lies at the root of all one’s choices, Kant refers to human beings as “radically evil.”

In laying out this account of radical evil, Kant clarifies important details about the nature of evil. For one thing, radical evil is not only “itself morally evil, since it must ultimately be sought in a free power of choice” (6: 37), but is also tied to a “natural propensity to evil” that structures particular evil choices that human beings make. Many commentators see this propensity to evil as a *precondition* of radical evil (e.g. Allison 1990, Wood 2000), but I see Kant as portraying the propensity to evil both as a *consequence* of humans’ radical evil and as a *ground* of further evil choices (see Frierson 2003). Moreover, the source of radical evil in choice implies that radical evil “cannot be placed, as is commonly done, in the sensuous nature of the human being and in the natural inclinations originating from it” (6: 34–35). For one thing, evil cannot be in the human being qua object of empirical investigation but must be traced to the free, noumenal agent that grounds empirically observable behavior. But even the empirical expression of radical evil is not in the lower faculties – the senses and inclinations – but in the higher faculties, especially in the higher faculty of desire. Human agents, even as empirically observed, have a capacity to act from principles, and the way this capacity is used gives empirical evidence of freely chosen evil.

Kant also describes three ways evil might express itself in one’s choices: frailty, impurity, and depravity. The first involves merely a lack of character, an “inability to act according to principle” (25: 650). Here the principles of one’s higher faculty of desire are good, but when it comes to acting, these principles do not actually determine one’s actions. As we noted in the last chapter, there can be conflicting underlying grounds of action, and often one or more powers are “dead” or “inactive” while others are active in effecting a transition to a new mental state or an action. Those with frail wills understand the principles according to which they should act, and the character of their higher faculty of desire is such that “I incorporate the good (law) into the maxim of my power of choice, but this good ... is subjectively the weaker (in comparison with inclination) whenever the maxim is to be followed” (6: 29). In the paradigm cases of frailty, one’s higher faculty of desire is properly oriented such that, if active, it would cause one to do what is right. But when the relevant moment comes, the higher faculty of desire is weaker than inclination (the lower faculty of desire) and hence inactive.

The other two forms of evil involve acting in accordance with corrupted principles. “Impurity” occurs when one’s “maxim is good with respect to its object ... [but] has not ... adopted the law alone as its sufficient incentive” (6: 30). One who is impure generally chooses what is morally required, but always only because it is both morally required and conducive to satisfying other desires. Such conditional adherence to the moral law is not real adherence. The final form of radical evil, “depravity,” involves a specific “propensity of the power of choice to maxims that subordinate the incentives of the moral law to others (not moral ones)” (6: 30). The depraved person might often act in seemingly moral ways, but his power of choice is structured by a fundamental commitment to non-moral desires, regardless of whether these are morally permitted or not.

Importantly, Kant rejects the possibility of what he calls “diabolical” evil, the “disposition ... to incorporate evil qua evil ... into one’s maxim” (6: 37). For Kant, even the most evil person is not motivated by evil as such. Thus Kant does not allow the possibility of cases like St. Augustine’s famous theft of pears “not to eat for ourselves, but simply to throw to the pigs[, where] our real pleasure consisted in doing something that was forbidden” (Augustine 1961: 47). For Kant, Augustine’s self-diagnosis must be mistaken; human beings do not have a desire to do what is morally forbidden *per se*. Evil arises only from putting non-moral desires ahead of our innate moral predisposition.

Finally, in all of these cases, radical evil need not imply that one always chooses contrary to the moral law. To be evil is to be disposed to allow the moral law to be overridden given a sufficient sensuous incentive. Frailty, impurity, and even depravity all involve, in different ways, a subordination of the moral law to non-moral desires. But one can be radically evil and still often do what is good, if one does what is good only because the price of doing good is, in a particular case, not too high (see 6: 39).

### *The problem of radical evil*

Kant’s claim that human beings are radically evil raises a serious problem at the intersection of transcendental and empirical anthropology, a problem that Kant spends the rest of his *Religion* trying to solve. Put simply, because this evil “corrupts the grounds of all maxims” it seems that it cannot “be extirpated through human forces, for this could happen only through good maxims – something that cannot take place if the subjective supreme ground of all maxims is presupposed to be corrupted” (6: 37). We cannot extirpate evil from our power of choice through that same (evil) power of choice. Radical evil is a consequence of humans’ use of their transcendental freedom. But given that we freely choose evil *as the basis of all of*



## 80 *Kant on the Human Being*

*our other choices*, it seems impossible to use that same freedom to rid ourselves of evil. The problem of radical evil is made even worse by the fact that human beings not only choose in evil ways but also cultivate themselves and their environment (especially their social environment) to promote the easy exercise of evil tendencies. Finally, the problem is even more acute because no matter how good one might be able to become, one has chosen badly, so one can never be a person who *always* chooses in accordance with the moral law (6: 72). Altogether, not only is one's choice oriented in such a way that one rejects moral reform (*radical* evil), but even if one were somehow to begin such a process of reform, one would have to contend with self-wrought influences that make morally upright action difficult (a *propensity* to evil), and even if one somehow overcame these influences, one would never have a life that was wholly good from start to finish (one *started from* evil).

Nonetheless, Kant defends moral hope, the possibility of reforming oneself morally despite one's radical evil. But this commitment to hope generates a problem: how can one reconcile moral rigorism, radical evil, and moral hope? At one level, Kant does not even try to explain how moral reform is possible given radical evil. He points out that evil cannot be extirpated "through human forces" (6: 37) and adds, "Some supernatural cooperation is also needed" (6: 44). This "supernatural cooperation" is ultimately beyond rational comprehension and even practical use (see 6: 117–18, 191; 7: 43–44). The main role of this "grace" is to reinforce humans' need to do their part to "make themselves antecedently worthy of receiving it" (6: 44). Kant emphasizes that the inscrutability of grace is no greater than the inscrutability of freedom and even that humans' continuing recognition of their moral obligations reveals an enduring "germ of goodness ... that cannot be extirpated or corrupted" (6: 45–46). The enduring germ of goodness shows that all people still have a capacity for goodness, and one's freedom gives an enduring but inexplicable hope that this capacity can still be used well. Of course, none of these claims about inscrutability actually address the central problem of radical evil.

But Kant's theoretically inadequate discussion of radical evil highlights the proper stance toward the problem.<sup>1</sup> Given his transcendental anthropology of cognition, Kant is correct that the metaphysical mechanisms by virtue of which radical evil might be overcome will never be understood by human beings. But the problem of radical evil is not, fundamentally, a metaphysical problem but a practical one. What ought one do in light of radical evil and what may one hope with respect to it? If evil is a free choice to subordinate the moral law to non-moral desires, one must simply subordinate non-moral desires to the moral law. But radical evil is also a self-wrought tendency to act immorally, and it is, moreover, a tendency evident in humans by nature. And these aspects of radical evil

require some grounds for moral hope in the human species as a whole as well as an account of how one can work to undo and arm oneself against self-wrought evil tendencies. Kant deals with the former task in his sophisticated philosophy of human history, a history situated in the context of radical evil but one that justifies hope in humanity's future. Kant deals with the second task in his "moral anthropology," which deals with "the subjective conditions in human nature that hinder people or help them in fulfilling [moral] laws ... , with the development, spreading, and strengthening of moral principles" (6: 217). The rest of this chapter focuses on Kant's philosophy of history. Kant's moral anthropology will be discussed in Chapter 5.

### **Human Beings as a Historical Species**

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

#### *Methodology*

Like the anthropology discussed in the last chapter, Kant's historical methodology is primarily empirical. Kant begins his essay "Idea for a Universal History with a Cosmopolitan Purpose" by emphasizing that "human actions" as "appearances ... are determined just as much as every other natural occurrence in accordance with universal laws of nature" and "History ... concerns itself with the narration of these appearances" (8: 17). But history is not "mere empirical groping without a guiding principle" (8: 161), and Kant's account of predispositions provides this principle. While the empirical anthropology of the previous chapter focused on predispositions as bases of causal powers, Kant's history studies predispositions teleologically. In his *Critique of Judgment*, Kant argued that organic life could be interpreted via purposive predispositions (5: 376). In writings on history, Kant adds that "all natural predispositions of a creature are determined sometime to develop themselves completely and purposively" (8: 18). For most animals, this teleological assumption has implications only for the study of individual organisms. To identify a feature of an