

For Agnes

RADICAL EVIL

A Philosophical Interrogation

Richard J. Bernstein

Radical Evil: Kant at War with Himself

It is inherent in our entire philosophic tradition that we cannot conceive of a "radical evil," and this is true both for Christian theology, which conceded even to the Devil himself a celestial origin, as well as for Kant, the only philosopher who, in the word he coined for it, at least must have suspected the existence of this evil even though he immediately rationalized it in the concept of a "perverted ill will," that could be explained by comprehensible motives. Therefore, we actually have nothing to fall back on in order to understand a phenomenon that nevertheless confronts us with its overpowering reality and breaks down all standards we know. . . . Totalitarian solutions may well survive the fall of totalitarian regimes in the form of strong temptations which will come up whenever it seems impossible to alleviate political, social, or economic misery in a manner worthy of man.

Hannah Arendt, *The Origins of Totalitarianism*

It was really as if an abyss had opened . . . *This ought not to have happened.* And I don't mean just the number of victims. I mean the method, the fabrication of corpses and so on – I don't need to go into that. This should not have happened. Something happened there, to which we cannot reconcile ourselves. None of us ever can.

Hannah Arendt, "What Remains? The Language Remains"

I have begun with these two epigraphs from Hannah Arendt because they help to orient my discussion of Kant. The first quotation is from the closing remarks of *The Origins of Totalitarianism*; and the second is from a television interview that she gave in 1964 in which she recalls her shock

when she first discovered what was taking place in Nazi death camps.¹ Ever since Kant used the expression “radical evil” (*radikal Böse*) in *Die Religion innerhalb der Grenzen der blossen Vernunft*, it has been a source of fascination and perplexity: fascination, because it has struck many of his readers (including Arendt) that Kant was dimly aware of a type of evil that exceeds our traditional conceptions of evil; perplexity, because it is not clear precisely what Kant means by “radical evil,” or how it fits (or does not) with his moral philosophy. I want to probe the meaning of “radical evil” for Kant, the philosophic context in which he explores its significance, and how radical evil is related to his moral philosophy. Is it true, as Arendt suggests, that Kant suspected the existence of a type of evil that calls into question our traditional ways of understanding evil? Is it true, as Arendt claims, that he “immediately rationalized it in the concept of a ‘perverted ill will’ that could be explained by comprehensible motives”?

It would be anachronistic to expect that Kant anticipated the horrors of the twentieth century. But Kant is certainly a thinker who has transformed the way in which we think about morality in the modern world. Despite the many critiques of Kant’s conception of morality, he has not only inspired subsequent thinkers, but we are presently living through a resurgence of interest in, and novel appropriations of, Kant’s moral philosophy. So it is eminently appropriate to ask whether Kant’s reflections on morality and radical evil can help orient our own thinking about the evil we have witnessed in the twentieth century. It is in this spirit that I approach Kant and interrogate him.

The primary analysis and discussion of radical evil is to be found in *Religion within the Limits of Reason Alone*.² The opening sentences of the preface reiterate Kant’s fundamental conviction that “for its own sake morality does not need religion at all.”

So far as morality is based upon the conception of man as a free agent who, just because he is free, binds himself through his reason to unconditioned laws, it stands in need neither of the idea of another Being over him, for him to apprehend his duty, nor of an incentive other than the law itself, for him to do his duty. At least it is man’s own fault if he is subject to such a need; and if he is, this need can be relieved through nothing outside himself: for whatever does not originate in himself and his own freedom in no way compensates for the deficiency of his morality. Hence for its own sake morality does not need religion at all (whether objectively, as regards willing, or subjectively, as regards ability [to act]); by virtue of pure practical reason it is self-sufficient. (*Rel.* 3; 3)

Kant not only asserts the independence of morality; but it is also clear from this passage that human beings are fully accountable and completely

responsible for what they do as free moral agents – whether they do their duty and obey the moral law or whether they fail to act in accordance with the moral law. If we are to understand what Kant means by radical evil, then our first task is to understand what Kant means by evil. Kant tells us that good or evil must “lie only in a rule made by the will [*Willkür*] for the use of its freedom, that is, in a maxim” (*Rel.* 17; 19). This is extremely important, for we will soon see that neither our natural inclinations nor our reason is the source of evil, but only our will. Just as Kant singles out the will as the primary locus of what is good, so it is the will that is the primary locus of evil. Consequently, we can say that good and evil have reference to the *maxims* of human volition.

We can already see how the *Religion* clarifies a troubling ambiguity in Kant’s moral philosophy. This concerns his understanding of the will. In the *Groundwork*, Kant appears to identify the will with practical reason. But if this were a strict identity, then how would it be possible for someone to commit an immoral or evil act? His younger contemporary, Reinhold, already criticized the identification of the will with practical reason during Kant’s lifetime, and this objection has been reiterated over and over again by many subsequent critics of Kant.³ But the *Religion* makes it eminently clear that Kant has a more complex and subtle understanding of the faculty of volition. This is indicated by the distinction that he introduces between *Wille* and *Willkür* (which unfortunately are frequently both translated as “will” in English).⁴ Although Kant is not always consistent, in general, when he refers to the will as the capacity to choose between alternatives, he calls it *Willkür*. The human *Willkür* (as distinguished from the *Willkür* of brute animals) is the faculty of free spontaneous choice. Or, more accurately, it is that aspect of the faculty of volition that involves unconstrained free choice. As Kant tells us, “the freedom of the will [*Willkür*] is of a wholly unique nature in that an incentive can determine the will [*Willkür*] to an action *only so far as the individual has incorporated it into his maxim* (has made it the general rule in accordance with which he will conduct himself); only thus can an incentive, whatever it may be, coexist with the absolute spontaneity of the will [*Willkür*] (i.e. freedom).” (*Rel.* 19; 23).⁵ The *Willkür*, the name we give to the capacity to choose between alternatives is neither *intrinsically* good nor *intrinsically* evil; rather, it is the capacity by which we freely choose good or evil maxims. In the *Religion* it is clear that *Wille* (in its more technical, narrow sense) does not act at all; it does not make decisions. *Wille* refers to the purely rational aspect of the faculty of volition. Henry Allison states the point succinctly when he writes: “Kant uses the terms *Wille* and *Willkür* to characterize respectively the legislative and executive functions of a unified faculty of volition, which he likewise refers to

as *Wille*.⁷⁶ And John Silber gives a lucid description of the relation of the *Wille* to the *Willkür*.

Unlike *Willkür*, however, *Wille* does not make decisions or adopt maxims; it does not act. Rather it is the source of a strong and ever present incentive in *Willkür*, and, if strong enough to be adopted by *Willkür* into the maxim of its choice, *Wille* "can determine the *Willkür*" and then "it is practical reason itself." *Wille* expresses the possibility of autonomy which is presupposed by transcendental freedom. The *Wille* represents the will's own demand for self-fulfillment by commanding *Willkür*, that aspect of the will which can either fulfill or abnegate its freedom, to actualize its free nature by willing in accordance with the law (and condition) of freedom. The most important difference between *Wille* and *Willkür* is apparent here. Whereas *Willkür* is free to actualize either the autonomous or heteronomous potentialities of transcendental freedom, *Wille* is not free at all. *Wille* is rather the law of freedom, the normative aspect of the will, which as norm is neither free nor unfree. Having no freedom of action, *Wille* is under no constraint or pressure. It exerts, instead, the pressure of its own normative rational nature upon the *Willkür*.⁷⁷

It is vital to see why Kant makes this all-important distinction between *Wille* and *Willkür*. When he introduces the categorical imperative in the *Groundwork*, we are left with an awkward consequence. If the will is completely identified with practical reason, then it is not clear where choice enters into making a moral decision. But Kant's account of morality presupposes that we are agents who have the capacity to choose freely to obey or disobey what is dictated by the moral law. Moral responsibility requires this capacity. To have the capacity to choose does not mean that we are indifferent. On the contrary, to the extent that we respond to what the *Wille* as the moral norm – as the law of freedom – dictates, we are autonomous. Furthermore, the moral law can be a *sufficient* incentive for us to do what duty requires. But, as Silber properly emphasizes, we must also be able to make heteronomous choices. Nothing determines the *Willkür* unless the *Willkür* chooses to be so determined. In this sense, to be a *finite* rational agent is to be radically free – that is, to be an agent who can choose good or evil maxims.⁷⁸

Evil maxims

But what precisely is an evil maxim? Before addressing this question directly, I want to emphasize that Kant is perfectly clear that it is not our natural inclinations that are the source of evil. On the contrary, he does

not even say that the existence of natural inclinations is neutral (neither good nor evil), but rather, that they are actually *good*! (Later we shall consider in what sense they are good.) Kant declares: "Natural inclinations, considered in themselves, are good, that is, not a matter of reproach, and it is not only futile to want to extirpate them but to do so would also be harmful and blameworthy. Rather, let them be tamed and instead of clashing with one another they can be brought into harmony in a wholeness which is called happiness." (*Rel.* 51; 60).

He explicitly repudiates the caricature that is so frequently drawn of him. He is frequently, but mistakenly, criticized for allegedly claiming that it is our natural inclinations that are the source of human evil. This caricature is misleading, because it obscures what is fundamental for his understanding of freedom and morality – that human beings, by virtue of their faculty of volition, are completely accountable and responsible for the good and evil maxims that they adopt. In *this respect*, there is no original sin or evil, just as there is no original moral goodness. To put the point positively, all sins, vices, and virtues originate in a (free) *Willkür*. The primary issue for Kant is always how we *choose* to respond to different, and sometimes conflicting, incentives.

Man *himself* must make or have made himself into whatever, in a moral sense, whether good or evil, he is or is to become. Either condition must be an effect of his [*Willkür*]; for otherwise he could not be held responsible for it and could therefore be *morally* neither good nor evil. When it is said, Man is created good, this can mean nothing more than: He is created *for* good and the original *predisposition* in man is good, not that, thereby, he is already actually good, but rather that he brings it about that he becomes good or evil, according to whether he adopts or does not adopt into his maxim the incentives which this predisposition carries with it ([an act] which must be left wholly to his own free choice). (*Rel.* 40; 48)

Yirmiyahu Yovel emphatically makes Kant's point: "Kant insisted in the *Religion* that evil too originates in freedom. This principle bars the view that when acting immorally we are causally determined by natural inclinations, and makes responsibility for evil possible. Nature cannot generate evil; only the free human will can."⁷⁸

In order to clarify the nature of duty and the moral law in the *Groundwork*, Kant focuses on those situations in which there is a conflict between our natural inclinations and our moral duty. In his famous (and sometimes misleading) examples, he tends to suggest that the paradigmatic examples of acting morally occur only when there is an overt clash between what we naturally desire and what we recognize as our duty, what we ought to do. This is the source of another persistent caricature of

Kant: we are only truly moral when we are acting *against* our natural inclinations. But here too, the *Religion* repudiates this caricature, and helps to clear up this misunderstanding. The basic issue for Kant in determining whether a maxim is good or evil is not whether it “contains” the incentive to follow the moral law *or* our natural inclinations. Rather, the issue is how these incentives are ordered – which incentive is primary and which one is secondary, that is, *subordinated*.

Hence the distinction between a good man and one who is evil cannot lie in the difference between the incentives which they adopt into their maxim (not in the content of the maxim), but rather must depend upon *subordination* (the form of the maxim) i.e., *which of the two incentives he makes the condition of the other*. Consequently man (even the best) is evil only in that he reverses the moral order of the incentives when he adopts them into his maxim. He adopts, indeed, the moral law along with the law of self-love; yet when he becomes aware that they cannot remain on a par with each other but that one must be subordinated to the other as a supreme condition, he makes the incentive of self-love and its inclinations the condition of obedience to the moral law; whereas on the contrary, the latter, as the *supreme condition* of the satisfaction of the former, ought to have been adopted into the universal maxim of the will [*Willkür*] as the sole incentive. (*Rel.* 31–2; 38)

As Allen Wood declares:

all maxims of finite rational volition (be they good or evil) contain *both* the incentives of moral reason and of sensible inclination; every maxim must contain both these incentives if it is to be the principle from which a finite rational subject acts, since both incentives belong to the predisposition of such a subject. . . . The maxim of the good man differs from that of the evil man only in that the former conditions the incentives of inclination by those of duty, whereas the latter reverses the moral order of incentives and makes it a rule to do his duty only on the condition that it be consistent with the pursuit of inclination.⁹

Consequently I can, and frequently do, act morally in a manner that is consonant with my sense of what is morally required and with what I naturally desire to do. I do not frustrate my natural inclinations. The moral worth of an agent depends exclusively on how these different incentives are ordered in the maxim that he adopts. But, of course, as Kant has noted already in the *Groundwork*, I may be mistaken about what really is my primary incentive, and whether a genuinely moral incentive has motivated the adoption of a maxim.

Although this understanding of the difference between a good and an evil maxim clears up a persistent misunderstanding about the role

of inclinations in our maxims and our actions, it does have some strong (Kant would say “rigoristic”) consequences. If we consider the example from the *Groundwork* of the shopkeeper who feigns honesty because this is the most advantageous and profitable policy, we can well understand how in his maxim there is an ordering such that his primary incentive is to maximize his profit rather than to do what is morally required. He may act in accord with duty, but not for the sake of duty. But consider the more difficult example of the person whose primary motivation is his sympathy for his fellow human beings. Kant tells us that there are persons “so sympathetically constituted that without any motive or vanity of selfishness they find an inner satisfaction in spreading joy and rejoice in the contentment of others which have made this possible.” Although actions performed on this basis are “dutiful and amiable” and “deserve praise and encouragement,” they do not evince moral worth, because they are not done from duty. Christine Korsgaard has given a very sensitive and insightful analysis of this example.¹⁰ She notes that Kant clearly distinguishes this example, where one acts “from direct inclination (perform an action because one enjoys it)” from the shopkeeper example, where one acts “from indirect action (performs an action as a means to an end).”¹¹ Explicating what Kant means, Korsgaard writes:

Therefore, when Kant says that the difference between the sympathetic person and the dutiful person rests in their maxims, the contrast he has in mind is this: although the sympathetic person and the dutiful person both have the purpose of helping others, they have adopted this purpose on different grounds. The sympathetic person sees helping as something pleasant, and that is why he makes it his end. The morally worthy person sees helping as something called for, or necessary, and this is what motivates him to make it his end.¹²

I think that this is exactly right, although Kant in the *Religion* gives another reason for placing greater moral weight on the person who acts for the sake of duty rather than out of the motivation of sympathy.

Consider the consequence of Korsgaard’s explication when we put it together with Kant’s analysis of good and evil maxims in the *Religion*. Korsgaard tells us that “Duty is not a different purpose, but a different ground for the adoption of a purpose. So Kant’s idea here is captured by saying that the sympathetic person’s motive is *shallower* than the morally worthy person’s: both want to help, but there is available a *further* stretch of motivating thought about helping which the merely sympathetic person has not engaged in.”¹³ Korsgaard’s interpretative suggestion is eminently

reasonable, but she ignores a consequence of Kant's rigorism. For she describes the motivation of "the merely sympathetic person" in a way that clearly indicates that the individual is giving priority to his sympathy (a natural inclination) rather than to what his moral obligation requires. And this is the paradigm of what Kant takes to be an *evil maxim*. Indeed, Kant himself explicitly makes this point in the *Religion*.

For when incentives other than the law itself (such as ambition, self-love in general, yes, even a kindly instinct [*gutherziger Instinkt*] such as sympathy) are necessary to determine the will [*Willkür*] to conduct *conformable to the law*, it is merely accidental that these causes coincide with the law, for they could equally well incite its violation. The maxim, then, in terms of whose goodness all moral worth of the individual must be appraised, is thus contrary to the law, and the man, despite all his good deeds, is nevertheless evil. (*Rel.* 26; 31)

The person who (self-consciously) gives priority to his sympathetic feeling in his maxim is not just "shallower" than the moral person; he is evil. He is adopting an evil maxim.¹⁴ Yet Kant does not flinch from drawing this conclusion. In the *Religion*, he endorses such a rigorist analysis. He tells us that we call a man evil not because he performs actions that are evil, "but because these actions are of such a nature that we may infer from them the presence in him of evil maxims" (*Rel.* 16; 18). Furthermore, Kant accepts the exclusive disjunction that "*Man is* (by nature) *either morally good or morally evil*" (*Rel.* 17; 18). Kant admits that although "experience actually seems to substantiate the middle ground between the two extremes," nevertheless, it is "of great consequence to ethics in general to avoid admitting, so long as it is possible, of anything morally intermediate. . . . Those who are partial to this strict mode of thinking are usually called *rigorists* (a name which is intended to carry reproach, but which actually praises)" (*Rel.* 18; 21). Whoever incorporates the moral law into his maxim and gives it priority is morally good; and whoever fails to do this, but gives priority to other nonmoral incentives (including sympathy) is morally evil. And to drive home his point, Kant affirms: "Neither can a man be morally good in some ways and at the same time morally evil in others. His being good in one way means that he has incorporated the moral law into his maxim; were he, therefore, at the same time evil in another way, while his maxim would be universal as based on the moral law of obedience to duty, which is essentially single and universal, it would at the same time be only particular; but this is a contradiction" (*Rel.* 20; 23). Allison provides a clear statement of Kant's grand either/or – his rigorism. He tells us:

Starting with the premise that respect for the law is an incentive, Kant reasons that since the freedom of the will (*Willkür*) entails that an incentive can determine the will only if it is "taken up" into a maxim, it follows that the failure to make it one's incentive, that is, the failure to make the thought of duty or respect for the law the sufficient motivation for one's conduct, must be regarded as resting on the adoption of an alternative principle of action. But since the adoption of an alternative principle involves an explicit deviation from the law, such an act must be characterized as "evil."¹⁵

So, following out the logic of Kant's rigorist analysis, there does not seem to be any way to avoid the conclusion that a benign sympathetic person (who gives the incentive of sympathy priority over the moral law in his maxim), Hitler, and even Eichmann (whose maxims presumably did not give priority to respect for the moral law) are *all* morally evil. Kant would certainly acknowledge that there are differences among them. Despite his "official" doctrine, he recognizes such differences. Nevertheless, given the exclusive rigorist disjunction – good or evil – we must judge them to be evil. What they have in common is "the failure to make the thought of duty or respect for the law the sufficient motivation for one's conduct."

Radical evil

Thus far, I have been addressing the question of what Kant means by evil, specifically what constitutes a morally evil maxim; but I have not yet said anything about radical evil. How does Kant's understanding of radical evil supplement what he has told us about evil maxims? More generally, we want to know how radical evil is compatible with his moral theory.¹⁶

Why did Kant (so late in his career) feel the need to introduce the concept of radical evil? There are several reasons. There is no doubt that he wanted to extract and defend what he took to be the moral rational core of Christian religious faith. But I also think that there is a deeper philosophical reason. Without compromising his moral stance that human beings are responsible for their good and evil maxims and deeds, Kant's understanding of human nature is that we are neither angels nor devils. He rejected the idea that we are born morally good and become corrupted, as well as the idea that we are intrinsically morally evil, that we are born sinners, and consequently cannot escape from actually sinning. His understanding of human nature as intrinsically neither morally good nor morally evil also has significant consequences for his understanding of human history and progress. Kant seeks to walk a fine line. On the one hand, he is skeptical of the idea of moral progress whereby human beings

can (and will) achieve human perfection. On the other hand, although human beings can never escape from the propensity to evil – a propensity constitutive of their species nature – there can be moral progress in history insofar as human beings can become actually good by virtue of their freedom. Kant's faith in (limited) moral and political progress is played out against a dark background, a realistic appraisal of "crooked humanity." In this respect, Kant departs significantly from some of the more naïve and optimistic Enlightenment conceptions of human progress (for example, Condorcet). Many of the tensions and problems in Kant's conception of radical evil can be traced back to his attempt to reconcile the claim that human beings are, by their very nature, evil with the claim that, despite this propensity to evil, human beings (even those who are wicked) can become morally good.

Despite the striking connotations of the term "radical," Kant is not speaking about a special *type* of evil or evil maxim. He would not agree with Arendt when she declares that radical evil is a phenomenon that "confronts us with its overpowering reality and breaks down all standards we know."¹⁷ And he certainly does not mean anything like what Arendt means when she claims that "radical evil has emerged in connection with a system [totalitarianism] in which all men have become equally superfluous."¹⁸ But what does Kant mean? And why does he introduce this tantalizing concept? The answer to this question is complex, and we need to introduce a number of important distinctions in order to answer it.

To set the context for our analysis, we must first consider briefly the sharp distinction that Kant makes in his Critical Philosophy between phenomena and noumena. This distinction, which is so central to Kant's understanding of human freedom, has proved especially troublesome. There are many passages in Kant that seem to suggest a "two-world" theory in which there is no (and can be no) interaction between the two. If this were really so – if phenomena and noumena referred to two "ontologically" different worlds – then there would be no way to make coherent sense of his moral philosophy. Recently, a number of sympathetic commentators have argued that there is a more adequate way of interpreting (and/or correcting) Kant which shows that he is not committed to a rigid ontological distinction between two entirely different realms or worlds. Although they approach the relevant issues in a variety of different ways, these commentators, who include Silber, Allison, Wood, and Korsgaard, present a strong case for the claim. Thus Silber: "Although he [Kant] asserted that the two realms exist 'independently of one another and without interfering with each other,' he found it impossible to speak of moral problems without presupposing their complete interaction. . . . The experience of moral obligation is a prime example of thorough interaction. If the same

human being (and, therefore, the same *Willkür*) were not both moral and natural, existing fully and simultaneously in both realms, moral experience would be impossible."¹⁹ Korsgaard argues that the confusions regarding the two-world theory stem from "a failure to appreciate the radical nature of Kant's separation of theoretical and practical reason, and of their respective domains of explanation and deliberation. When these domains are separated in the way that Kant's philosophy requires, the problems about responsibility disappear, and we see that Kant's theory of freedom does not commit him to an ontological dualism."²⁰ Although these commentators have not cleared up all the problems that arise from the distinction between noumena and phenomena, I do think that they have been successful in showing that Kant is not committed to an extreme ontological dualism, and that Kant does have a unified conception of the human agent who is both free and conditioned by natural causality.

In the *Religion*, Kant does not abandon the noumena/phenomena distinction, but he plays down its significance. In the preface to the second edition, he says:

To understand this book in its essential content, only common morality is needed, without meddling with the *Critique of Practical Reason*, still less with theoretical Critique. When, for example, virtue as skill in actions conforming to duty (according to their legality) is called *virtus phaenomenon*, and the same virtue as an enduring disposition towards such actions from duty (because of their morality) is called *virtus noumenon*, these expressions are used only because of the schools; while the matter itself is contained, though in other words, in the most popular children's instructions and sermons, and is easily understood. (*Rel.* 12–13; 15–16)

The reason why it is so important to see that Kant is not committed to a two-world theory is because his analysis of radical evil presupposes the intelligibility of speaking about a "human nature" which cannot be identified simply with our phenomenal nature (or with our noumenal selves).

"Nature" and "human nature" are used in the *Religion* in a manner that is strikingly different from the typical use of these terms in the *Critique of Pure Reason* and the *Critique of Practical Reason*. Human nature encompasses what we are as phenomenal and moral beings. This becomes evident when, for example, Kant tells us that human nature possesses three predispositions [*Anlagen*] to good. To be more precise, these are three divisions, or elements "in the fixed character and destiny [*Bestimmung*] of man: (1) the predisposition to *animality* in man, taken as a *living* being; (2) the predisposition to *humanity* in man, taken as a *living* and at the same time a *rational* being; (3) the predisposition to *personality* in man, taken as a rational and at the same time an *accountable* being" (*Rel.* 21; 25). It is

perfectly clear from this passage that Kant is speaking about "human nature" (*der menschlichen Natur*) in a way that encompasses what he has previously classified as phenomena and noumena. When Kant speaks of "man" here, he is not referring to individual men. He is referring to man – or rather human beings – as the human race or species. He tells us, "The man of whom we say, 'He is by nature good or evil,' is to be understood not as a single individual (for then one man could be considered as good, by nature, another as evil), but as the entire race" (*Rel.* 21; 24–5). We must be careful not to misinterpret what Kant is affirming. We may be tempted to think that these three divisions of the original predisposition to good are constituents of a predisposition to be *morally* good. This temptation may be especially strong in the case of the predisposition to personality, because this is "the capacity for respect for the moral law as *in itself* a sufficient incentive of the will [*Willkür*]" (*Rel.* 23; 27). But the capacity for respect is not to be identified with the *actual* (moral) exercise of this capacity by the *Willkür*. We, as human beings, may have a *predisposition* to become morally good, but it is only by exercising our free will (*Willkür*) that we *actually* become morally good (or evil).

Man *himself* must make or have made himself into whatever, in a moral sense, whether good or evil, he is or is to become. Either condition must be an effect of his free choice [*Willkür*]; for otherwise he could not be held responsible for it and could therefore be *morally* neither good nor evil. When it is said, Man is created good, this can mean nothing more than: He is created *for good* and the original *predisposition* is good; not that, thereby, he is already actually good, but rather he brings it about that he becomes good or evil, according to whether he adopts or does not adopt into his maxim the incentives which this predisposition carries with it ([an act] which must be left wholly to his own free choice). (*Rel.* 40; 48)

But if this is so, then in what sense can we say that the original predisposition is a predisposition to and for *good*? Wood is helpful in clarifying what Kant means.

These predispositions all belong to "human nature" in the sense that they are "bound up with the possibility of human nature." They are predispositions *to good* in the sense that they are considered in themselves "not a matter of reproach" and that through them man is created "for good."...

No man is *actually* good or evil on account of his possession of these predispositions. Hence, if man is to be said to be "by nature" good or evil, this goodness or evil cannot consist in the predispositions bound up with the possibility of human nature. The very concept of morally good and evil involves, rather the actual use man makes of his capacities, and prevents us from regarding these capacities themselves as morally good or evil.²¹

It would have been clearer if Kant had simply said that although human beings have a predisposition to become morally good, they are not actually born morally good. They become morally good only if they freely choose to act so that they incorporate the moral law into their maxims.

Before turning explicitly to what Kant means by radical evil, we must first clarify what he means by disposition (*Gesinnung*) and propensity (*Hang*). For they are the basis for understanding radical evil. These concepts are not to be confused with what he calls a predisposition (*Anlage*).²² Silber claims that "the development of [the concept of *Gesinnung*] is, perhaps, the most important single contribution of the *Religion* to Kant's ethical theory, for by means of it he accounts for the continuity and responsibility in the free exercise of *Willkür* and for the possibility of ambivalent volition, as well as the basis for its complex assessment."²³ We can appreciate why Silber makes this strong claim. On the basis of the *Groundwork* and the *Critique of Practical Reason*, it is not entirely clear how Kant deals with the continuity of moral agency. In part, this is because he has focused primarily on the role of maxims in making moral decisions. But the individual who adopts maxims and makes choices is not simply a collection of discrete choices and maxims. Nor is he simply a timeless noumenal self. As Silber perceptively notes, "The disposition [*Gesinnung*] is thus the enduring aspect of *Willkür*, it is *Willkür* considered in terms of the continuity and fullness of its free expression. It is the enduring pattern of intention that can be inferred from the many discrete acts of *Willkür* and reveals their ultimate motive." And he adds: "Continuity in disposition [*Gesinnung*] is essential to moral self-identity."²⁴ This helps to clarify the work that the concept of *Gesinnung* is intended to perform, and why it is so essential to Kant's analysis of moral agency.

As long as we stay at this general level of abstraction, we do not encounter any difficulties. But the more closely we examine the details of what Kant says about *Gesinnung*, the more problematic this concept becomes. Initially we might think that Kant is finally coming to acknowledge the importance of what Aristotle recognized long ago, that he means something that closely approximates what Aristotle called *hexis*, an acquired disposition to act virtuously, a disposition that requires a proper upbringing and education. But Kant explicitly says that "this disposition [*Gesinnung*] itself must have been adopted by free choice [*Willkür*]" (*Rel.* 20; 24). The passage from which this comes is even more perplexing, because Kant says that, although this disposition is acquired, it "has not been acquired in time."

To have a good or evil disposition [*Gesinnung*] as an inborn natural constitution does not here mean that it has not been acquired by the man who harbors it, that he is not the author of it, but rather, that it has not been

acquired in time (that he has *always* been good, or evil *from his youth up*). The disposition, *i.e.*, the ultimate subjective ground of the adoption of maxims, can be one only and applies universally to the whole use of freedom. Yet this disposition itself must have been adopted by free choice [*Willkür*], for otherwise it could not be imputed. (*Rel.* 20; 24)

This is an extremely perplexing and obscure passage, and we will have to keep it in mind as we proceed. But, for all its obscurity, there is no ambiguity regarding what Kant is saying about a good or evil disposition (*Gesinnung*) being “adopted by free choice.” This, perhaps, is the most significant difference between a predisposition (*Anlage*) and a disposition (*Gesinnung*); a *Gesinnung* is adopted by free choice, but an *Anlage* is *not* chosen; it is a constituent of human nature, and it is bound up “with the possibility of human nature.” This freely chosen character of *Gesinnung* has led one critic to suggest that it is an unstable combination of Aristotelian *hexis* and a Sartrean “*projet fondamental*.”²⁵ We can see how radically Kant’s notion of a *Gesinnung* departs from any ordinary or traditional concept of a disposition when he tells us that “it has not been acquired in time,” and that this *Gesinnung* must itself be conceived of as a maxim – a supreme maxim that orients the moral life of an agent viewed as a whole (even though it is possible to alter this disposition).²⁶ One might think that here Kant comes close to being incoherent. He seems to get himself ensnared in a vicious circle: a good or evil disposition (*Gesinnung*) is adopted by free choice (*Willkür*), but *Willkür* itself presupposes this disposition.

Without underestimating these problems, I do think that we can give an account of *Gesinnung* that at least *approximates* what Kant wants to say (and is compatible with much of what he actually does say). For the moment, let us set aside the problems arising from the noumena/phenomena distinction and from what is and is not “acquired in time.” We frequently do distinguish persons with good and evil dispositions. There are persons whom we can trust or count on to do the right thing (what the moral law requires) in difficult moral situations. They exhibit an overall pattern of commitments, beliefs, intentions, and actions that provides a basis for our (fallible) judgment that when they confront difficult moral situations, they will do what duty requires. Or, to phrase this point in a Kantian manner, they will adopt good maxims, maxims that incorporate and give priority to the moral law. Of course, there are persons who are so selfish or narcissistic that they are likely to do anything to achieve their ends and to avoid acting for the sake of duty. There is no difficulty in drawing a distinction between a person’s disposition or overall moral character and the specific maxims he adopts and the actions he performs. Although his pattern of action is the basis for making judgments about his

character, we recognize that it is this character that informs his specific choices. The relation between moral character and the specific actions performed is neither strictly causal nor strictly logical. A person with a good disposition may occasionally act out of character and adopt a maxim that subordinates the moral law to natural inclination. And a moral scoundrel may occasionally do what duty requires (honor among thieves).

Viewed in this manner, we can understand why such a disposition may be characterized as the subjective ground for the adoption of specific maxims. For such a disposition *informs*, but does not (causally) determine the maxims that are adopted in specific situations of moral conflict. We can also understand the sense in which we are responsible for our dispositions or characters. It is not that at a specific moment in time we “choose” this disposition. Nevertheless, our disposition is the result of free moral decisions that we make and the maxims that we adopt. We are not born morally good or evil; we *become* morally good or evil by virtue of the choices we make.

If we press further and ask why one person develops a disposition or character that leads him to adopt good maxims and someone else adopts evil maxims, there is much that we can say about their background, social circumstances, and education; but we cannot give an ultimate answer to this question: it is inscrutable. For such an answer would require us to be able to give a *theoretical* account of human freedom. And this is precisely what the Critical Philosophy shows us to be impossible. To claim that a free choice is inscrutable is not to say that it is *mysterious* – as if, in principle, we should be able to give necessary and sufficient reasons for why someone makes the choices he does make; it is only to insist that the choice is *free*. Ultimately, we cannot *know* why one person chooses to follow the moral law and another person does not. Nevertheless, from a *practical* point of view, we can (and must) postulate freedom, and assert that moral agents have the capacity to choose freely good or evil maxims. We are responsible for our individual choices and for our overall moral character.

Furthermore, this disposition is not something fixed and unchangeable. A good person can become evil, and an evil person can become good. “For man, therefore, who despite a corrupted heart yet possesses a good will (*Wille*), there remains the hope of a return to the good from which he has strayed” (*Rel.* 39; 47). We can even demystify what seems so counter-intuitive, that a *Gesinnung* “has not been acquired in time.” We have to distinguish *two* different senses of being acquired in time. Kant is certainly not denying the obvious fact that we commonly specify the *time* when a moral decision is made. All of the examples that Kant gives in the *Groundwork* presuppose such a temporal location. We make our moral decisions

in concrete situations at specific times and places. For example, *tomorrow*, I may be faced with a terrible moral dilemma in which I will be confronted with a choice about what I ought to do. But there is a more *technical* sense of something happening in time that is explored in the *First Critique* when Kant gives his analysis of causality. *In this sense*, to say that something happens in time is to say that it is causally determined, and consequently not free. When Kant tells us that a *Gesinnung* is not acquired in time, he is telling us that it is not (naturally) causally determined but, rather, issues from our freedom.

I do not want to suggest that this understanding of *Gesinnung* clears up all the problems and difficulties that Kant's discussion involves; but it does help to show the plausibility of the idea of a *Gesinnung*, the sense in which it is a disposition, and yet freely chosen, and why such a notion is so important for a robust understanding of human agency. There is, however, one *major* difficulty with my analysis that cannot be avoided. Ironically, this points up one of the most troubling features of Kant's conception of radical evil. If my description approximates what Kant means by a *Gesinnung*, then it is applicable to *both* good and evil dispositions, or characters. And most of the time, Kant writes in a manner that would lead us to think that there are good and bad dispositions. Yet, when Kant turns explicitly to the notion of radical evil, and characterizes it as a propensity (*Hang*), he neglects the symmetry between good and evil dispositions. Most of the time he writes as if there is no significant difference between a disposition (*Gesinnung*) and a propensity (*Hang*), but he *never* speaks about a propensity to good, but only a propensity to evil.²⁷ He explicitly tells us that "man is evil by nature" (although we will see that this does not mean quite what we might think it does). Kant also insists that although "the character (good or evil) distinguishing man from other possible rational beings . . . is innate in him. . . we shall ever take the position that nature is not to bear the blame (if it is evil) or take credit for it (if it is good), but that man himself is its author" (Rel. 17; 20). In short, what Kant calls a propensity (*Hang*) cannot be identified with a disposition (*Gesinnung*), yet he fails to clarify what distinguishes these two concepts. He never explains why the disposition (*Gesinnung*) of human beings can be good or evil, whereas there is a propensity (*Hang*) *only* to evil.

Although there are certainly many obscurities and difficulties in Kant's characterization of *Gesinnung*, we can begin to appreciate why Kant gets himself entangled in these difficulties. They spring from one of the most admirable and central features of his moral theory. The many caricatures of him, notwithstanding, Kant certainly acknowledges that temperament, background, social circumstances, moral training (and even geographical location) of persons influence their moral character and choices. But the

primary issue for Kant is always the individual's accountability and responsibility. Our moral freedom is never compromised by external events or by our natural inclinations. We can make Kant's point even more forcefully by asserting that we alone are ultimately responsible for the individual moral choices that we make and for our moral character. Kant even seems to be aware of the awkward way in which he has described a *Gesinnung*. For he characterizes it as "the ultimate subjective ground of the adoption of maxims," and then immediately adds, "But the subjective ground or cause of this adoption cannot further be known (though it is inevitable that we should inquire into it)" (Rel. 20; 24).

We are finally in a position to turn explicitly to the concept of radical evil. On the basis of what we have learned about "the original predisposition [*Anlage*] to good in human nature," we at least know what radical evil is not. Radical evil is not, in any way, to be identified with radical inclinations. Radical evil is not to be identified with our phenomenal sensuous nature. The body, with its needs and desires, is not the source of evil. Nor is radical evil to be identified with any *intrinsic* defect or corruption of human reason. It is related solely to the corruption of the will. We can locate an essential clue about the meaning of radical evil, and why Kant introduces this concept, by returning to his analysis of the third division of the original predisposition (*Anlage*) to good that is inherent in our human nature. Let us consider again what Kant says about the predisposition to *personality*.

The predisposition to *personality* is the capacity for respect for the moral law as *in itself a sufficient incentive of the will* [*Willkür*]. This capacity for simple respect for the moral law within us would thus be moral feeling, which in and through itself does not constitute an end of the natural predisposition except so far as it is the motivating force of the will [*Willkür*]. Since this is possible only when the free will [*Willkür*] incorporates such moral feeling into its maxim, the property of such a will [*Willkür*] is good character. The latter, like every character of the free will [*Willkür*], is something that can only be acquired; its possibility, however, demands the presence in our nature of a predisposition on which it is absolutely impossible to graft anything evil. We cannot rightly call the idea of the moral law, with the respect which is inseparable from it, a *predisposition to personality*; it is personality itself. . . . But the subjective ground for the adoption into our maxims of this respect as a motivating force seems to be an adjunct to our personality, and thus to deserve the name of a predisposition to its furtherance. (Rel. 22-3; 27)

There is nothing startlingly new here that we could not have inferred from the *Groundwork* or the *Critique of Practical Reason*. But if this predispo-

sition to good is constitutive of our very nature as human beings, how are we to account for the fact that we do not always follow this predisposition; that we do not always do what we morally ought to do? Human beings are tempted to disregard the moral law, to adopt evil maxims – maxims that give priority to nonmoral incentives. It is this tendency or propensity (*Hang*) that Kant seeks to isolate with the introduction of the concept of radical evil. But what does Kant mean by a propensity (*Hang*)? He tells us that “by propensity (*propensio*) I understand the subjective ground of the possibility of an inclination (habitual craving, *concupiscentia*) so far as mankind in general is liable to it” (*Rel.* 23–4; 28). And how are we to distinguish a propensity (*Hang*) from a predisposition (*Anlage*)? “A propensity is distinguished from a predisposition by the fact that although it can indeed be innate, it *ought* not to be represented merely thus; for it can also be regarded as having been *acquired* (if it is good), or *brought by man upon himself* (if it is evil)” (*Rel.* 24; 28–9). We get a clearer idea of what Kant means by turning to his shocking footnote. He informs us that:

A propensity (*Hang*) is really only the *predisposition* . . . to crave a delight which, when once experienced, arouses in the subject an *inclination* to it. Thus all savage peoples have a propensity for intoxicants; for though many of them are wholly ignorant of intoxication and in consequence have absolutely no craving for an intoxicant, let them but once sample it and there is aroused in them an almost inextinguishable craving for it. (*Rel.* 24; 28)³⁸

“Radical evil” then, is not the name of some special type of evil (as Arendt maintains). And it certainly is not a form of evil that “we cannot conceive.” On the contrary, we can clearly conceive it, and what it names is the propensity (*Hang*) not to do what duty requires, not to follow the moral law. Indeed, Kant’s purpose in using the adjective *radikal* to qualify *Böse* is to indicate that this propensity is *rooted* in human nature, specifically in the corruption of the will (*Willkür*); he is appealing to the original, etymological meaning of *radikal*. There is no evidence that Kant means anything more than this.

Kant does distinguish different degrees of the “capacity for evil,” but all three are related to the failure to adopt good maxims.

First, there is the weakness of the human heart in the general observance of adopted maxims, or in other words, the *frailty* of human nature; second, the propensity for mixing unmoral with moral motivating causes (even when it is done with good intent and under maxims of the good), that is, *impurity*; third, the propensity to adopt evil maxims, that is, the *wickedness* of human nature, or of the human heart. (*Rel.* 24; 29)

We may think that “wickedness” names some horrendous form of evil. And Kant’s rhetoric certainly makes it sound this way.

Third: the wickedness (*vitiositas, pravitas*) or, if you like, the *corruption* (*corruptio*) of the human heart is the propensity of the will [*Willkür*] to maxims which neglect the incentives springing from the moral law in favor of others which are not moral. It may also be called the *perversity* (*perversitas*) of the human heart, for it reverses the ethical order [of priority] among the incentives of a *free will* [*Willkür*]; and although conduct which is lawfully good (i.e., legal) may be found with it, yet the cast of mind is thereby corrupted at its root (so far as the moral disposition is concerned), and the man is hence designated as evil.

It will be remarked that this propensity to evil is here ascribed (as regards conduct) to men in general, even to the best of them; this must be the case if it is to be proved that the propensity to evil in mankind is universal, or, what here comes to the same thing, that it is woven into human nature. (*Rel.* 25; 30)

Let us reflect carefully on Kant’s characterization of the “third degree” of the capacity for evil with reference to the example of the sympathetic person that I discussed earlier. Suppose that such a person – even when it is pointed out to him – makes a conscious choice to continue to give priority to his feelings of sympathy for his fellow human beings as the *primary* incentive for his maxims (rather than “incentives springing from the moral law”). He trusts his heart more than his reason, even though his heart may occasionally lead him astray. And let us also grant, as Kant says we may, that the conduct springing from the incentive of sympathy is “lawfully good.” He consistently acts in accord with duty, although not for the sake of duty. On the basis of Kant’s classification of the degrees of evil, we would be compelled to judge such person as *wicked*. He presumably has a cast of mind that is corrupted at its root. Why? Because he gives priority to what Kant considers to be a nonmoral maximum when he ought to adopt a moral one. And even though, by hypothesis, his actions conform to the moral law, this is simply accidental, or contingent. But to judge such a person as the paradigm of *wickedness*, to put him in the same category as a mass murderer – at least, in respect to the degree of evil exhibited – is more than an awkward consequence of Kant’s rigorism; it is morally perverse.

But we face still greater problems. Kant’s analysis of radical evil as a propensity that is “woven into human nature” actually *obscures* (rather than clarifies) a cardinal point in his moral philosophy. The very concept of a propensity (*Hang*) is one that is parasitic upon our notion of causality. A propensity presumably has causal efficacy. Thus, in Kant’s unfortunate

example of a savage people who crave intoxicants, we think of such a propensity as having overwhelming causal power. Such a craving demands satisfaction unless it is resisted in the strongest possible manner. But the propensity to evil cannot be thought of in this way. It is not an active causal force "pushing" us, or tempting us, to be morally evil. There is no moral evil unless we *freely* adopt evil maxims. A *Willkür* that adopts such maxims is not causally determined by anything but itself; it is the spontaneous manifestation of our freedom. Few philosophers have been as insistent as Kant has been, in arguing that genuine freedom is unconditioned by any (natural) causal influences.²⁹

It may be objected that the example of the craving for intoxicants is misleading, because it is a physical propensity, and Kant himself makes a sharp distinction between a physical (natural) propensity and a moral propensity.

Every propensity is either physical, i.e., pertaining to the will [*Willkür*] of man as a natural being, or moral, i.e., pertaining to his will [*Willkür*] as a moral being. In the first sense there is no propensity to moral evil, for such a propensity must spring from freedom; and a physical propensity (grounded in sensuous impulses) towards any use of freedom whatsoever – whether for good or bad – is a contradiction. Hence a propensity to evil can inhere only in the moral capacity of the will [*Willkür*]. But nothing is morally evil (i.e., capable of being imputed) but that which is our own *act*. (Rel. 26; 31)

Kant realizes that the very idea of a moral propensity as "a subjective ground of the will [*Willkür*] which precedes all acts" is a problematic notion. If such a propensity results from the exercise of freedom, then this propensity must itself issue from an act of free will (*Willkür*). Although it may seem *ad hoc* and a bit contrived, Kant introduces *two* senses of "act" to resolve this problem. In the first sense, "act" refers to the exercise of freedom whereby the *Willkür* adopts the supreme maxim. (This is what Yovel calls a "global moral strategy.") The second sense refers to specific acts performed on the basis of this supreme maxim. But these distinctions do not alter the main point that I want to emphasize; on the contrary, they reinforce it. The alleged physical propensity for intoxicants is neither universal nor necessary; it is not a propensity of human beings as a species. It need not even result in the adoption of evil maxims. One can exercise one's free will (*Willkür*) to resist this temptation. If a propensity to moral evil "springs from freedom," one may begin to wonder whether there really is such a propensity. Why? Because if the propensity "springs from freedom," then its very existence depends upon "that exercise of freedom whereby the supreme maxim . . . is adopted by the will [*Willkür*]"

(Rel. 26; 31). But the *Willkür* is the capacity for choosing maxims freely. Choosing the supreme maxim, the subjective determining ground, is itself an act of the *Willkür*. But the *Willkür* is not conditioned or causally-influenced by any propensity, physical or moral.

We can examine Kant's problem from a slightly different perspective. Kant makes two claims which, although not necessarily incompatible, nevertheless seem to undermine the very idea of a moral propensity to evil. The first is that this propensity is itself the result of an *act* understood as the exercise of freedom whereby the supreme maxim is adopted by the will (*Willkür*). The second is that we, as free moral agents, can always resist this alleged propensity (which we have adopted by the exercise of our freedom). But if both these claims are true, then it is difficult to understand what is left of the very idea of a "propensity to moral evil." It is extraordinarily paradoxical (if not incoherent) to claim that there is a propensity to moral evil that is universal and, "as it were, rooted in humanity itself," and yet that "we must, after all, ever hold man himself responsible for it" (Rel. 28; 33). Yet this is precisely what Kant does maintain. He unambiguously affirms that this is what we call "radical innate evil in human nature (yet none the less brought upon us by ourselves)" (Rel. 28; 33). He says that "We must not, however, look for an origin in time of a moral character [*Beschaffenheit*] for which we are held responsible; though to do so is inevitable if we wish to explain the contingent existence of this character" (Rel. 38; 46).

Readers might want to counter my reading of Kant by noting that when he introduces the idea of a propensity (*Hang*), he emphasizes that it is a *possibility*. He says, "By propensity (*propensio*) I understand the subjective ground of the possibility of an inclination . . ." (*Unter einem Hange* (propensio) verstehe ich den subjektiven Grund der Möglichkeit einer Neigung) (Rel. 23; 28). Consequently, there is no incompatibility in ascribing such a propensity to human nature and affirming that human beings have the capacity of free choice (*Willkür*). But this is not the source of the difficulty that I find in Kant. Stressing possibility (*Möglichkeit*) does not distinguish a propensity from a predisposition. A predisposition (*Anlage*) is also "bound up with the possibility of human nature" (Rel. 23; 28). The main problem concerns the origin or source of this propensity to evil. Kant insists that we are the authors of this propensity; that it results from the exercise of *our* freedom; that it is "brought by man upon himself" (Rel. 24; 29). This is what is so difficult to accept: namely, that the propensity to evil is innate or inborn (*angeboren*), yet we are somehow responsible for it.

Sometimes, we can detect what appear to be opposing and contradictory claims within a single sentence. Consider one of the most famous (and frequently quoted passages) from the *Religion*: "This evil is radical,

because it corrupts the ground of all maxims; it is, moreover, as a natural propensity, *inextirpable* by human powers, since extirpation could occur only through good maxims, and cannot take place when the ultimate subjective ground of all maxims is postulated as corrupt" (Rel. 32; 39). If the sentence had ended there, then a straightforward reading would lead us to think that evil is radical because the will (*Willkür*) is corrupt at its very source or origin. This is a very strong claim indeed, and can readily be assimilated to a secular version of the Christian doctrine of original sin. (This is the reading of radical evil that seems to have offended Goethe.) But consider how Kant completes the sentence: "yet at the same time it must be possible to *overcome* it, since it is found in man, a being whose actions are free" [*gleichwohl aber muss er zu überwiegen möglich sein, weil er in dem Menschen als frei handelndem Wesen angetroffen wird*] (Rel. 32; 39).

Suppose we ask, how is it possible to overcome this inextirpable natural propensity? The answer for Kant *must* be that this occurs by a free exercise of the will (*Willkür*). For this is the way in which "a being whose actions are free" manifests his freedom. But then, contrary to what has been affirmed in the first part of the sentence, it cannot be true that radical evil corrupts the ground of *all* maxims. If it did, there would be *no* possibility of adopting or willing a good maxim, and consequently no possibility of overcoming radical evil. The claim that Kant makes in the second part of the sentence is something he affirms over and over again in the *Religion*. Indeed, he makes an even more forceful claim. All human beings are radically evil – that is, possess the powerful propensity to become morally evil – but only some persons *do* become morally evil, and develop a morally evil character or disposition. But even such wicked persons can be reborn and become good. "A change of heart . . . must be possible because duty requires it" (Rel. 60; 70).

It may be thought that we can come to Kant's rescue by making a sharp distinction between human beings as a *species* and *individual* human beings. There is no doubt that when Kant affirms that man by nature is evil, he is referring to the human species. But even though he affirms that the species has a propensity to evil, he is not saying that this propensity determines the moral character of individual persons. Individuals may develop good or evil dispositions by virtue of the free choices they make. But this distinction between species and individuals does not really help to clarify the issue. On the contrary, it leads to greater difficulties. For if the propensity to evil "springs from freedom," and this propensity, and it is ascribable to the human species, then we would have to say that the human species *qua* species freely chooses this propensity. It is not clear that such a thesis is even intelligible.

Why does Kant allow himself to get entangled in such difficulties and

paradoxes? It looks as if he wants to have his cake and eat it too! And in a way he does. Or, to switch metaphors, Kant is at war with himself. For, on the one hand, he never wants to compromise the basic claim of his moral philosophy: that human beings as finite rational agents are free, which means that they are *solely* and *completely* responsible for their moral choices and for the maxims they adopt. If we become morally good or evil, this is our own doing and a consequence of our own free will (*Willkür*). On the other hand, Kant also wants to affirm that all human beings have an innate propensity to moral evil. In order to have his cake and eat it too, he is then driven to claim that even though this propensity is woven into the fabric of human nature, it is a propensity that springs from our freedom, and one for which we are responsible. Later, I want to show that what at first seems at best an extremely strained and awkward position, and at worst a blatant contradiction, actually reveals one of the most enduring and attractive features of Kant's moral philosophy.

The more we focus on the details of Kant's analysis of radical evil, the more innocuous the concept seems to be (despite Kant's rhetoric about human wickedness).³⁰ After making the apparently dramatic claim that "man is evil by nature," Kant goes on to say, "Man is *evil*, can mean only, he is conscious of the moral law but has nevertheless adopted into his maxim the (occasional) deviation therefrom" (Rel. 27; 33). But do we need the *Religion* or any special concept of radical evil to know this? The *Groundwork* – indeed, the very project of Kant's moral philosophy – is based upon the idea that we do not always do what we ought to do; that we, as finite rational agents, are *not* holy wills, and consequently do not always follow the moral law. Presumably, the introduction of the concept of radical evil is intended to explain *why* (from a practical point of view) we deviate from following the moral law. We do not always follow the moral law *because*, as human beings, we have an innate propensity to evil. Our wills are corrupted at their root. But does this "because" really explain anything? Does it do any conceptual work? I do not think so. When stripped down to bare essentials, it simply reiterates the fact that human beings who are conscious of the moral law sometimes (freely) deviate from it. Furthermore, it is *always* within our power to resist this propensity, no matter how strong it is supposed to be. In short, radical evil – the alleged propensity to moral evil which is a universal characteristic of human beings – does not have *any* explanatory force (practical or theoretical) at all!

I have no doubt that Kant intended to make a much more forceful claim, that he thought he was showing something really fundamental about human beings when he asserted that man by nature is evil. My argument is that there is a disparity between what he *intends* and what he

ways. When we scrutinize what he actually says, when we see how he qualifies his key claims, it is difficult to avoid the conclusion that Kant himself eviscerates the notion of radical evil.

We have not yet come to the end of our difficulties with the concept of radical evil. According to Kant, radical evil is a *species* concept; it is universally applicable to all human beings woven into the very fabric of human nature. We all have an "evil heart." It is not, then, a contingent characteristic of *some* human beings, or even a contingent characteristic of *all* human beings. But what is the *justification* for making such a bold and controversial claim? If there is one lesson that we should have learned from the Critical Philosophy, it is that genuinely synthetic universal claims can never be justified by appeal to experience; their justification requires a "deduction" – a proof. Yet, when Kant reaches this crucial stage in his exposition, when we expect some sort of proof or justification of radical evil as a *universal* characteristic of human beings, *no* such proof is forthcoming. This is what Kant says: "That such a corrupt propensity must indeed be rooted in man need not be formally proved in view of the multitude of crying examples which experience of the actions of men puts before our eyes" (*Rel.* 28; 33–4). Kant follows this assertion with some empirical observations based upon (dubious) anthropological evidence; "melancholy" observations about "civilized peoples" and casual remarks about the nefarious international behavior of nation-states (*Rel.* 28–9; 34).³¹ Henry Allison states the serious problem we confront here quite clearly, and even attempts to do what Kant himself failed to do: to provide an a priori deduction that will justify the claim that there is a universal propensity to moral evil.

Kant insists not only that there is a propensity to evil but that it is "rooted in humanity itself" and, therefore, universal. What grounds, we may ask, does Kant offer for this apparently audacious claim?

Kant's official answer to this obvious question is quite disappointing. . . . Instead of offering a "formal proof" of the universality of the propensity to evil, he simply asserts the necessity for such a proof is obviated by "the multitude of crying examples which experience of the actions of men put before our eyes." In short, he seems to treat it as an unproblematic empirical generalization. But clearly, even if for the sake of argument one accepts Kant's appeal to some rather selective anthropological evidence, the *most* that this evidence can show is that evil is widespread, not that there is a universal propensity to it. Moreover, since Kant insists that this propensity concerns only the ultimate subjective ground of one's maxims and is perfectly compatible with a virtuous empirical character, it is difficult to see what could conceivably falsify this claim. Consequently, it is also difficult to take seriously the suggestion that it is intended as an empirical generalization.³²

Kant never gives – or even attempts to give – a *proof* of his controversial and bold claim that man is evil by nature.

It almost appears as if Kant is caught in what he has characterized (in the *Critique of Pure Reason*) as a "dialectical illusion." Such an illusion arises when we think we have genuine knowledge and can explain something, but actually do not, and cannot, have any legitimate knowledge. Starting from the fact that human beings sometimes adopt good maxims and sometimes adopt evil maxims, we seek to explain why they do not always follow the moral law. We presumably *explain* this failure by appealing to the doctrine of radical evil – the propensity towards moral evil that is rooted in our humanity. But it is an illusion to think that this enables us to explain or account for why we adopt evil maxims, why we sometimes succumb to this temptation. This alleged explanation turns out to be vacuous. For it does not explain anything about the individual free choices that we make, or even why we choose the ultimate subjective ground of our maxims. To ask *why* we freely choose to adopt good or evil maxims is to ask an *impossible question*. It is an impossible question because, ultimately, it is "inscrutable to us."

When we say, then, Man is by nature good, or Man is by nature evil, this means only that there is in him an ultimate ground (inscrutable to us [*uns uneiferschlüchlen*]) of the adoption of good maxims or of evil maxims (i.e., those contrary to law) and this he has, being a man; and hence he thereby expresses the character of his species. (*Rel.* 17; 20)

To drive home the point that the adoption of the ultimate subjective ground is inscrutable, Kant adds the following important footnote:

That the ultimate subjective ground of the adoption of moral maxims is inscrutable is indeed already evident from this, that since this adoption is free, its ground (why, for example, I have chosen an evil and not a good maxim) must not be sought in any natural impulse, but always again in a maxim. Now since this maxim also must have its ground, and since apart from maxims *no determining ground* of free choice [*Willkür*] can or ought to be adduced, we are referred back endlessly in the series of subjective determining grounds, without ever being able to reach the ultimate ground. (*Rel.* 17–18; 20)

It appears, then, that the concept of radical evil is a dialectical illusion because it seduces us into thinking that we can *explain* something that we cannot possibly explain – why we freely adopt the maxims (good or evil) that we actually do adopt – whether it be the choice of an ultimate subjective ground of maxims or the choice of specific maxims in concrete situations.

Diabolical evil

My aim in this chapter has been to understand what Kant means by radical evil. But, in the spirit of interrogation, I also want to explore how Kant's reflections about morality and radical evil help us to understand the forms of evil that have broken out in the twentieth century, including the evil and genocide evoked by the word "Auschwitz." John Silber (and many would agree with him) says that Kant is "the most important writer on ethics since Aristotle," and that he intended to "set forth doctrines that were not just theoretical in nature but were intended to guide everyday human conduct."³³ Despite Kant's language of "wickedness," "corruption," and "perversity," his analysis of evil and radical evil is disappointing. But some of Kant's reflections on duty – especially the absolute duty of a citizen to obey the sovereign power, and the duty of a soldier to obey orders of a superior – are more than disappointing; they are extremely disturbing.

Consider the case of Adolf Eichmann, who cited Kant (with reasonable accuracy) at his trial in Jerusalem in order to justify his conduct.³⁴ We certainly cannot blame Kant, the great champion of human dignity, for this perverse appropriation of the categorical imperative. Nevertheless, as Silber points out, "It may seem outrageous to find Kant's ethical doctrine, grounded as it is in the dignity of the moral person as an end-in-himself, used to exculpate a confessed accomplice to mass murder. But it should come as no surprise to Kant scholars, for Kant's views on the citizen's obligation to the sovereign strongly support Eichmann's position."³⁵ Kant's official doctrine is that the ban on resisting any supreme lawmaking is *absolute*. (Let us not forget that, for all the manipulation, lawlessness, and violence of the Nazis, "Hitler was made Chancellor in a constitutionally proper manner."³⁶) I want to quote just three of the many passages that Eichmann might have cited from Kant to justify his acceptance of the *Führerprinzip*.

Any resistance to the supreme lawmaking power, any incitement of dissatisfied subjects to action, any uprising that bursts into rebellion – that all is the worst, most punishable crime in a community. For it shatters the community's foundations. And this ban is *absolute*, so unconditional that even though that supreme power or its agent, the head of state, may have broken the original contract, even though in the subject's eyes he may have forfeited the right to legislate by empowering the government to rule tyrannically by sheer violence, even then the subject is allowed no resistance, no violent counteraction.

There is no right of sedition (*seditio*), much less a right of revolution (*rebellio*), and least of all a right to lay hands on or take the life of the chief of state

when he is an individual person on the excuse that he has misused his authority. . . . It is the people's duty to endure even the most intolerable abuse of supreme authority.³⁷

Thus it would be ruinous if an officer, receiving an order from his superiors, wanted while on duty to engage openly in subtle reasoning about its appropriateness or utility; he must obey.³⁸

I certainly do not want to judge Kant by the way in which his statements have been misappropriated and distorted. Nor is it fair to criticize him for a failure to anticipate the systematic terror and violence practiced by the Nazis. I have no doubt that Kant, the great champion of universal human dignity, would have found ample grounds to condemn the Nazis. Nevertheless, the consistency, and even harshness, with which Kant opposes any active resistance to "the supreme lawmaking power," no matter how tyrannical it may become, should at least make us question his rigorism his insistence that this ban is *absolute* and *unconditional*.³⁹

The Eichmann question concerns a soldier's duty to obey his superiors, but what about the supreme commander, Hitler himself? How does Kant's moral theory apply to the person who is ultimately responsible for giving the orders? Does Kant's understanding of evil and radical evil help us to judge the conduct of Hitler?⁴⁰ I cannot explore the complex historical debate concerning Hitler's intentions and motivations. I refer to Hitler primarily in order to raise some further questions about Kant's understanding of evil that I have not yet squarely addressed. To raise these questions, we need to return to some of the details of Kant's analysis of radical evil. Let us recall that in the *Religion*, Kant makes it perfectly clear that neither our sensuous nature nor our faculty of reason is the source of evil. The locus of evil is the will – or, more precisely, the corruption of the will (*Willkür*). In this respect, Kant stands in a tradition that goes back to St Augustine. We, and we alone, are responsible for the evil maxims that we freely adopt by an act of will (*Willkür*). In a famous passage, Kant categorically rejects the possibility of thinking of man as "a devilish being" (*enim leyflischen Wesen*).

In seeking, therefore, a ground of the morally evil in man [we find that] *sensuous nature* comprises too little, for when the incentives which can spring from freedom are taken away, man is reduced to merely *animal* being. On the other hand, a reason exempt from the moral law, a *malignant reason* as it were (a thoroughly evil will [*Wille*]) comprises too much, for thereby opposition to the law would itself be set up as an incentive (since in the absence of incentives the will [*Willkür*] cannot be determined), and thus the subject would be made a *devilish* being. Neither of these designations is applicable to man. (*Rel.* 30: 37)

But why can't human beings be devilish beings? And why is the idea of a "malignant reason" rejected? Why does Kant simply rule this out as impossible? The search for answers to these questions takes us to the very heart of Kant's moral philosophy, and to his understanding of radical evil. There are reasons why Kant rejects these possibilities. Before we can evaluate this rejection, we need to understand his reasons.

Kant is primarily concerned with man as a species, with the human race. If a malignant reason were *constitutive* of human nature, if man as a species were intrinsically devilish, then there would be no morality. The reason is clear. Morality presupposes freedom and choice. If one claims that malignant reason is *constitutive* of our human nature, there is no possibility that we could act otherwise. This means that a moral agent, by the exercise of his will, has the capacity to choose freely the maxims he adopts.

What about the possibility that *some* human beings are devilish beings? Once again, if this is interpreted to mean that some but not all human beings are intrinsically devilish, then the same considerations apply. For this would mean that some human beings are not really human — they do not have the capacity to choose between good and evil maxims. But suppose we consider the case of someone who is not innately (in the strong sense of innate) diabolical, but who *becomes* diabolical — who freely and consistently chooses to defy the moral law. Kant addresses this possibility when he writes:

Man (even the most wicked) does not, under any maxim whatsoever, repudiate the moral law in the manner of a rebel (renouncing obedience to it).

The law, rather, forces itself upon him irresistibly by virtue of his moral predisposition; and were no other incentive working in opposition, he would adopt the law into his supreme maxim as the sufficient determining ground of his will [*Willkür*]; that is, he would be morally good. (*Rel.* 31; 37)

But this passage is strikingly ambiguous; it is open to at least two very different interpretations, which must be carefully distinguished. In order to bring out this ambiguity, a review of the *Wille/Willkür* distinction is in order. In the *Religion*, Kant emphasizes that the *Willkür* is that aspect of the faculty of volition by which we make free choices. Even when we recognize the moral law as the norm to which our maxims ought to conform, we nevertheless have the capacity to do (or not to do) what the moral law requires. We have the capacity to choose good or evil maxims. The *Wille* itself (in the narrow technical sense) is practical reason; it is the moral law, the supreme moral norm. The *Wille/Willkür* distinction is introduced within a *unified* faculty of volition (sometimes also referred to as *Wille*). In short, there is no *Wille* without a *Willkür*, and there is no (hu-

man) *Willkür* without the *Wille*. *Wille* and *Willkür* are co-dependent, although we can distinguish their different functions. Allison crisply states the difference when he says: "Thus, it is *Wille* in the narrow sense that provides the norm and *Willkür* that chooses in light of this norm."⁴¹ When Kant says that "the law . . . forces itself upon [man] irresistibly," he is making not a *causal* claim but a *normative* moral claim; he is asserting that we (finite moral agents) cannot help but recognize the objective moral law as the norm to which our maxims ought to conform. We cannot help but *acknowledge* the categorical imperative, regardless of whether we choose to obey it or not. This is the sense in which it is perfectly accurate to declare: "Man (even the most wicked) does not, under any maxim whatsoever, repudiate the moral law in the manner of a rebel (renouncing obedience to it)." To be a human being is to be a person who *recognizes* the authority of the moral law regardless of whether one chooses to do what it requires.⁴²

We can now locate the crucial ambiguity in the above passage where Kant affirms the impossibility of man being a devilish being. Even if we accept his claim that human beings as a *species* are not devilish, and that no matter how wicked a person may be, he cannot avoid acknowledging the authority of the moral law, this does not address the issue of whether an *individual* can repudiate the moral law in the sense of freely choosing to defy it. I want to argue that this is not only possible, but also that, on Kant's own analysis of *Willkür*, it *must* be possible. It must be possible for an individual to *become* a devilish person. It must be possible for an individual to defy and repudiate the moral law in such a manner that he freely adopts a disposition (*Gesinnung*) in which he consistently refuses to do what the moral law requires. He consistently adopts evil maxims. This may be judged to be morally wicked and perverse, but nevertheless it is a perverse possibility.

To bring out the full significance of what I am claiming, I want to examine Silber's fundamental criticism of Kant's refusal to acknowledge the possibility of a human being becoming a "devilish being." Silber thinks that Kant's refusal to consider the possibility that individuals may consistently defy the moral law reveals a fundamental weakness in his moral philosophy. This is the primary reason why Silber claims that Kant's ethics is not adequate to account for Auschwitz.

Kant's ethics is inadequate to the understanding of Auschwitz because Kant denies the possibility of the deliberate rejection of the moral law. Not even a wicked man, Kant holds, can will evil for the sake of evil. His evil, according to Kant, consists merely in his willingness to ignore or subordinate the moral law when it interferes with his nonmoral but natural inclina-

tions. His evil is expressed in abandoning the conditions of free personal fulfillment in favor of fulfillment as a creature of natural desire. . . . [Kant denies] the possibility of a person knowingly doing evil for its own sake. By insisting that freedom is a power whose fulfillment depends upon rationality and that its irrational misuse is merely an impotence, Kant proposed a theory that rules out the contravening evidence of human experience.⁴³

Several commentators have sought to defend Kant against Silber's objection. Allen Wood, for example, addressing himself to an earlier version of this criticism by Silber, says: "This, however, is a fallacy endemic to philosophical criticism: the supposition that by pointing to 'facts' (which no one disputes) one can give a philosophical justification of the manner in which one has expressed the facts."⁴⁴ It is, of course, true that philosophical claims rarely are resolved by "pointing to 'facts,'" and that the crucial issue frequently turns out to be the *interpretation* of the alleged facts. Nevertheless, Wood's dismissal of Silber's objection strikes me as a bit too facile, for two reasons. In the first place, Kant himself — as we have seen — supports his own thesis about the universality of radical evil by "pointing to 'facts.'" Secondly, this is not quite what Silber means when he says, "Kant's insistence to the contrary, man's free power to reject the law in defiance is an ineradicable fact of human experience."⁴⁵ Silber is not referring to empirical "facts" and "experience" in the restricted technical sense in which these terms are used in the *Critique of Pure Reason*, where they refer to the phenomenal realm. He is using them more broadly, in a manner consistent with Kant's own usage when he speaks of "the *experience* [*Erfahrung*] of the actions of men" in the *Religion* (*Rel.* 28; 34). Silber claims that there are persons who deliberately and consistently reject the moral law, even though they recognize what the moral law requires, and that Kant does not adequately account for this possibility.⁴⁶

Silber might have been more precise had he phrased his point in a slightly different manner. An individual is judged to be evil if he chooses a disposition (*Gesinnung*) — a supreme, overarching maxim — to adopt evil maxims. We do not "observe" maxims directly; we infer their existence on the basis of actions that human beings perform. Such an inference is always fallible. We can never be absolutely certain when we ascribe a maxim (even a supreme maxim) to someone. But, despite this fallibility, we do judge persons to be evil (or good) on the basis of their actions. In this sense, our *experience* of the actions of human beings is the basis for making judgments about their evil character. So Silber might have said that, on the basis of our moral experience, we judge some persons to be devilish. And this is a possibility that Kant failed to consider.

We still haven't come to the heart of the matter — to what Silber is

getting at in his criticism of Kant. For Silber is not arguing that human beings as a species are devilish, or that being devilish is somehow constitutive of our nature. And he is not even denying that human beings must acknowledge the authority of the moral law. On the contrary, his criticism depends on affirming this claim. Rather, he is underscoring something that Kant does not seem to consider: namely, that there are some persons (as well as characters in fiction) who, to use the Kantian terminology, incorporate into their maxims the primary incentive to *defy* the moral law.⁴⁷

Although I agree with Silber that Kant does not explicitly deal with this possibility, ironically, his moral theory — as developed in the *Religion* — can effectively deal with it. In a very revealing footnote, Kant writes:

For from the fact that a being has reason it by no means follows that this reason, by the mere representing of the fitness of its maxims to be laid down as universal laws, is thereby rendered capable of determining the will [*Willkür*] unconditionally, so as to be "practical" of itself; at least not so far as we can see. The most rational mortal being in the world might still stand in need of certain incentives, originating in objects of desire, to determine his choice [*Willkür*]. He might, indeed, bestow the most rational reflection on all that concerns not only the greatest sum of these incentives in him but also the means of attaining the end thereby determined, without ever suspecting the possibility of such a thing as the absolutely imperative moral law which proclaims that it is itself an incentive, and, indeed, the highest. Were it not given us from within, we should never by any ratiocination subtilize it into existence to win over our will [*Willkür*] to it; yet this law is the only law which informs us of the independence of our will [*Willkür*] from determination by all other incentives (of our freedom) and at the same time of the accountability of all our actions. (*Rel.* 21; 25)

Although Kant's phrasing is a bit turgid, his basic point is clear. Reason by itself may not be sufficient to motivate us to follow the moral law. We may recognize the fitness of our maxims as laid down by the universal law, but we may nevertheless, not yet be motivated to adopt these maxims and act accordingly. "The most rational mortal being in the world might still stand in need of certain incentives . . . to determine his choice." Of course, our respect for the law *may be* a sufficient incentive to act morally. But our *Willkür* may choose to defy the moral law. If recognition of the moral law can serve as an incentive to act morally, there can always be a counter-incentive. We can choose to be perverse, we can choose to be devilish, we can choose to defy the moral law. We may be told that such a choice is irrational, that we are refusing to recognize "the absolutely imperative moral law," that there is a performative contradiction whereby

we are both exercising and denying our freedom. But it does not follow that we *cannot* do this! On the contrary, such a possibility is intrinsic to the human *Willkür*. There are no *intrinsic* restraints on what the *Willkür* can choose to do; we are "radically free."⁴⁸

We can approach the issue of radical free choice [*Willkür*] from a slightly different angle. Kant typically limits the incentives involved in the adoption of maxims to *two* kinds: the moral incentive to conform to the moral law and nonmoral incentives that arise from our natural inclinations and desires. At times, Kant even categorizes all nonmoral incentives under the rubric of self-love, "which, when taken as the principle of all our maxims, is the very source of evil" (*Rel.* 41; 49).⁴⁹ But why should we limit incentives to these two kinds? Why not recognize that there are other incentives that are not easily assimilated to "self-love." It is difficult to see how the incentives that motivate fanatics and terrorists who are willing to sacrifice themselves for some cause or movement can be accounted for by self-love. The horrors of the twentieth century (and not just this century) have opened our eyes to the variety of types of incentives that motivate evil actions.⁵⁰ Sometimes it seems as if Kant is operating with a highly abstract formal principle. If an incentive is not a genuinely moral incentive — that is, respect for the moral law — then it *must* (by stipulative fiat) be classified as the incentive of self-love.⁵¹ The difficulty here is rooted in Kant's limited moral psychology, in the narrow range of types of incentives that he acknowledges.⁵² If one is really to distinguish different types of evil, then one must consider the full range of incentives that are involved in the adoption of evil maxims and the performance of evil deeds. There are major differences among those who may be misguided because they give priority to their sympathetic feeling for their fellow human beings, those (like Eichmann) whose primary incentive for performing their "duty" seems to be advancing their own career, those who mock and defy the moral law, and those who do evil for evil's sake. I am *not* suggesting that Kant is unaware of these differences. There is plenty of evidence that he acknowledges them. But I am questioning whether he has provided the *conceptual* resources to account and illuminate them. To claim that *all* evil maxims are determined by the principle of self-love (no matter how broadly we think of self-love) obscures more than it illuminates. *There is no free choice (Willkür) unless there is the free choice to be morally evil, and even devilish.*

Unconditional moral responsibility

In concluding this chapter, I want to return to the question that motivated this analysis originally: whether Kant's reflections on evil, especially radi-

cal evil, can help to guide our thinking about the evils we have witnessed in the twentieth century. The answer is a mixed one, for we need to distinguish the strengths and weaknesses of his reflections. Kant would not have agreed with Arendt that "radical evil" names a special type of evil that cannot be conceived. I have argued that Kant's concept of radical evil turns out to be little more than a way of designating the tendency (propensity) of human beings to disobey the moral law. There is an enormous disparity between Kant's rhetoric — his references to "wickedness," "perversity," and "corruption" — and the content of what he actually says. Against Kant's explicit rejection of the possibility that man is (or can become) devilish or diabolical, I have argued that his understanding of *Willkür* entails that *some* individuals can *become* devilish; and this conclusion is a *necessary* consequence of Kant's understanding of free choice (*Willkür*).

I have also claimed that Kant is at war with himself, and I want to spell this out in a bit more detail. When we understand *why* Kant gets entangled in these difficulties and double binds, we begin to appreciate both his importance and his relevance to the attempt to come to grips with the problem of evil. There is one cardinal principle that Kant refuses to compromise in any way. This is at once the source of his difficulties and also his profound insight into morality. Human beings are morally accountable and responsible for whatever they become, for the maxims that they adopt, even for their moral disposition. Kant never compromises on the principle that it is *always* within our power to choose between good and evil maxims, and that it is we (and we alone) who must bear the responsibility for these choices. There are no moral excuses such that we can say that we have been *compelled by natural causes* to choose or to will what is morally evil. This is why, no matter how much Kant insists that radical evil is a powerful propensity or tendency, that it is innate, that it is inextirpable, he never interprets this to mean that we are *causally* compelled to choose evil maxims and to do evil deeds. He absolutely insists that no matter how deeply rooted this propensity is within our human nature, it is not this propensity that is responsible for the evil that we do, but our free will (*Willkür*). This is why every time we think Kant is telling us that our will is fundamentally corrupt, that we are evil by nature, that this evil is woven into the very fabric of our humanity, he immediately qualifies what he says, reminding us that we, and we alone, are responsible for what we do. There is no escape from the radical freedom of our *Willkür*. And radical freedom means we must bear the complete moral responsibility for our choices, decisions, and actions.

We can now see why Kant is so relevant for coming to grips with the many faces of evil in the twentieth century. His uncompromising insistence that personal responsibility is inescapable goes against the grain of

prevailing tendencies to find all sorts of excuses for our moral failures. Kant would sharply oppose the variety of “functional explanations” that seek to diminish the significance of individual responsibility, just as he would object to fashionable attempts to decenter or dissolve the moral agent such that it no longer even makes sense to speak about individual responsibility. Kant’s understanding of freedom – not only the freedom manifested in self-legislation and obeying the moral law, but the more radical freedom of choice (*Willkür*) that this moral freedom presupposes – enables us to evaluate the actions of individuals in extreme situations.

Let us return to the opening epigraphs of this chapter where Hannah Arendt speaks of radical evil. One of the most troubling issues that arises in the attempt to comprehend totalitarianism and the phenomenon epitomized by Auschwitz is the assignment of responsibility – not only to the perpetrators – to those who gave orders and those who followed orders – but also to the so-called bystanders. We do not have to say that all those involved are responsible in the same way. There are crucial moral and legal differences to be made between, for example, a Hitler, a Heydrich, an Eichmann, and those bystanders who actively or passively supported the Nazis. Even the victims had to make drastic choices. Kant would never have endorsed a notion of collective responsibility that entailed saying that an entire people were *equally* responsible. But he would have insisted that insofar as individuals have the capacity of spontaneous free choice (*Willkür*), they are accountable and responsible moral agents.⁵³

Finally, I want to consider again Kant’s claim that the “ultimate subjective ground of the adoption of moral maxims” is *inscrutable*. I have already quoted the passage from the beginning of the first essay of the *Religion* where he asserts “that the ultimate subjective ground of the adoption of moral maxims is inscrutable.” Kant not only begins his essay on radical evil with this claim, but he concludes the essay by reiterating that “the depths of the heart (the subjective first ground of his maxims) are inscrutable to him” (*Rel.* 46; 56). Some commentators see this claim about inscrutability as a problem, or, more ungenerously, as a “cop-out.” But I take it to be an indication of Kant’s ultimate intellectual integrity and his profound understanding of our radical freedom. When confronted with the moral choices that human beings make, there is a great deal that we can know about why persons make the choices they do. Kant was acutely aware of how background, cultural conditions, and education can affect moral choices. But we still want to know ultimately why one person chooses good maxims and another chooses evil maxims. Or, to stick close to Kant’s terminology, we want to know how we are to account for “the ultimate subjective ground of the adoption of moral maxims.” For this ultimate subjective ground must itself be the result of an exercise of free-

dom. In the final analysis, we cannot explain why one person chooses to become good and another chooses to become evil. To imagine that we could explain this would be in effect to deny that our will (*Willkür*) is radically free. So, far from its being some sort of deficiency, it is Kant’s way of acknowledging a profound moral truth about our radical free choice (*Willkür*). Human beings are responsible for the choices they make, but *ultimately*, we cannot explain why they make the moral choices they do; we cannot explain “the ultimate subjective ground of the adoption of moral maxims” – whether for good or for evil. Not only is this inscrutable; it *must* be inscrutable, because this is what it means to be a free and responsible person.