Evil and Imputation in Kant's Ethics

Mark Timmons

For Kant, moral evil of all sorts — evil that is rooted in a person's character — is manifested in action which, on the one hand, is explicable in terms of an agent's own reasons for action and so imputable, though on the other hand it is, in some sense, irrational. Because such evil is rooted in a person's character, it "corrupts the ground of all maxims"[1] and thus deserves to be called *radical* evil. Moreover, according to Kant, not only are human beings susceptible to such evil, being evil is an inescapable condition of being human. These claims raise a number of questions, among them the following: (1) How can we explain the possibility of irrational, yet explicable, freely done actions given Kant's views about human agency? (2) What is the nature of radical evil? (3) In what sense is it a corrupting ground of all maxims? (4) What reason does Kant have for claiming that radical evil is an inescapable part of the human condition? There are other questions to be added to this list, some of them addressed in the recent secondary literature, but for the most part I plan to focus on the ones just mentioned.

Specifically, my plan goes as follows. Sections I and II are concerned with some basic themes and distinctions pertaining to Kant's theory of action and practical rationality that will provide a basis for understanding his view of moral evil. In section III, I turn to Kant's analysis of moral evil in its various manifestations in order to provide, in section IV, a general characterization of moral evil. In section V, I consider Kant's claim that radical evil corrupts the ground of all maxims, and what this claim implies about the possibility of actions having moral worth. In section VI, I want to consider what we can call Kant's Universality

---

[1] Die Religion innerhalb der Grenzen der Blossen Vernunft VI: 37, 32. All references to Kant's works cite the volume and page number of the Prussian Akademie edition of Kant's works, (Kant's gesammelte Schriften, herausgegben von der Königlich Preußischen Akademie der Wissenschaften, 29 volumes, Berlin, Walter de Gryter & Co. 1902 ff.) followed by the page number of the English translations. The translated works I have used are: The Groundwork of the Metaphysics of Morals (Grundlegung zur Metaphysik der Sitten), translated by *H. J. Paton* (third edition) New York: Harper & Row, 1965; The Metaphysics of Morals (Die Metaphysik der Sitten), translated by *Mary Gregor,* Cambridge: Cambridge University Press, 1991; Critique of Practical Reason (Kritik der praktischen Vernunft), translated by *Lewis White Beck,* Indianapolis / New York: Bobbs-Merrill, 1956; and Religion within the Limits of Reason Alone (Die Religion innerhalb der Grenzen der bloßen Vernunft), translated by *Theodore M. Greene* and *Hoyt W. Hudson,* New York: Harper & Row, 1960.

Thesis, viz., his claim that necessarily all human beings are radically evil. What is remarkable about the Universality Thesis (UT) is Kant's apparent argument for it: despite the fact that this thesis supposedly holds with strict universality for all human beings and thus would appear to be a necessary truth, Kant defends it on empirical grounds. In his recent book, Henry Allison has argued that the UT represents a synthetic a priori claim, and he proceeds to provide a 'deduction' for it, something that Kant failed to do. I argue that Allison's deduction does not work, though I am afraid I don't have a deduction of my own to propose on Kant's behalf, nor do I see how to reconcile Kant's own defense of the UT with its alleged a priori status. Kant's UT remains problematic.

I. Maxims, Reasons, and Motivation

I interpret Kantian maxims as intentions of an agent which can be properly expressed by statements beginning 'I will. . . .' [2] They concern actions to be done or not done or ends to be achieved, and are adopted by agents for reasons, though due to such factors as ignorance and error on the agent's part, the act of adopting them may fail to be rational. Kant distinguishes between the 'matter'

[2] Two comments are in order here. First, my claim that maxims are intentions is controversial. *John E. Atwell,* Ends and Principles in Kant's Moral Thought, Dordrecht / Boston / Lancaster: Martinus Nijhoff, 1986, pp. 50-51 argues that maxims are not intentions. *Onora O'Neill,* "Consistency in Action", in: *Nelson T. Potter* and *Mark Timmons,* eds: Morality and Universality, Dordrecht: D. Reidel, 1985, pp. 161-7 argues that maxims are *fundamental* intentions — very general intentions that guide agents to accept more specific intentions on the basis of which they perform actions. Space does not permit me to defend my view, but see *Barbara Herman,* Morality as Rationality, New York: Garland, 1990, ch. 2, and *Henry E. Allison,* Kant's Theory of Freedom, Cambridge: Cambridge University Press, 1990, pp. 91-4 for a defense of the view I hold. Second, formulations of maxims may include more or less information about the agent's intentions. Focusing, for a moment on maxims of action, what we might call a *simple maxim* has the form: 'I will _____ if / whenever _____', where the first blank is filled with a specification of the action and the second with a specification of the circumstances under which the agent intends to perform the action. However, in Metaphysik der Sitten VI: 385, 189, Kant claims that all actions have ends and what we might call a *complex maxim* has the form: 'I will _____ if / whenever _____, in order to _____', where the first two blanks are filled as before and the last one mentions whatever end(s) the agent takes this action to serve. I construe complex maxims as a fusion of two distinct maxims: a simple maxim of action plus a maxim of ends. In attempting to understand why an agent performed some action, we often require that the action be related to the agent's ends, and a complex maxim, as I am calling it, expresses this relation. It should be noted that the end mentioned in a complex maxim may or may not represent the agent's most fundamental motive (motivating reason) for adopting a maxim of action. After all, I might intend to enroll in a series of cooking courses with the immediate end of learning fine French cuisine, though my more long range goal (and so what motivates my having this immediate end) is to start my own catering business. One could express this plan as a complex maxim having this form: 'I will _____, if / whenever _____, in order to _____, so that I might bring about _____.'

of maxims and their 'form.' The matter of a maxim is its content, and concerns that action (or end) to be adopted or avoided and the circumstances under which that action (end) is to be performed (pursued) or avoided. In the rest of this section, I want to focus on the 'form' of maxims by considering a number of themes and theses that are central in Kant's theory of practical reasoning.

1. Motivating Versus Justifying Reasons for Action

Since talk about reasons in connection with intentional action is ambiguous, we can avoid confusion if we make the following distinctions. First, let us call any explanation of an agent's action that makes sense of the action *as something the agent did* (as opposed to something that happened to the agent) a *rationalizing explanation.* Rationalizing explanations are explanations in terms of a person's reasons for action, and so explain actions from the agent's perspective. Second, let us call *motivating reasons* those features of the agent's situation that figure in a rationalizing explanation of her actions. In many cases, perhaps typical ones, there are features of some action and her situation that the agent takes to be a good or adequate reason for her performing that action. For example, if an agent is thirsty and believes that the glass before her contains thirst-quenching liquid, then her reason for drinking the liquid is the complex of her thirst together with her belief about the liquid.

But here, my talk of motivating reasons is intended quite broadly to cover not only reasons for action that the agent takes to be good or at least adequate reasons for action, but is also meant to include those factors, such as certain desires, emotions or passions that prompt an agent to do or refrain from certain courses of action, but which the agent does not take to be good or adequate reasons for action. For example, a person with a moral conscience but who has sadistic desires or urges may not take those desires or urges to provide any sort of reason for engaging in the sorts of actions those desires prompt. Again, someone addicted to tobacco and wanting desperately to quit may not take her craving cigarettes, or the enjoyment derived from smoking, to provide her with a reason for smoking. In these cases, we explain the actions of the sadist and the smoker by appealing to motivating considerations — the sadist's strong desire to inflict pain on his victim, and the smoker's craving for a cigarette — though such considerations are not taken by these individuals to be good or adequate reasons for action. They are however motivating reasons, at least in the sense that they are the sorts of factors that figure in rationalizing explanations of people's behavior. We can call motivating reasons for action that the agent takes to be good or adequate reasons, *endorsed reasons,* and those reasons that the agent does not take to be good or adequate reasons, *non-endorsed reasons.*

In addition to motivating reasons, there are *justifying reasons,* i. e., those considerations that really are good reasons for an agent to adopt this or that maxim and act accordingly. I say more about justifying reasons in the next section.

2. Two Sources of Reasons for Action

For Kant, there are two ultimate sources of reasons (both motivating and justifying) for action: one's own happiness and dictates of morality. Happiness is understood here as the satisfaction of one's desires — desires that arise ultimately from one's sensuous nature.[3] We may call those reasons for action that stem from considerations of one's own happiness, *desire-based reasons for action*. But, according to Kant, there is a source of reason for action that stems from our nature as rational creatures. That is, for Kant, some reasons for action depend on there being certain substantive principles of rationality, notably, moral principles. Such moral principles, or categorical imperatives, can be understood as principles of rationality telling an agent what maxims it is rational to adopt and act on. Reasons for action stemming from reason itself, we may call *reason-based reasons for action*.

In providing rationalizing explanations of a person's action, then, we ultimately appeal to desire-based reasons, non-desire-based reasons, or, in cases of overdetermination, reasons of both sorts.

3. The Incorporation Thesis

As I have already mentioned, according to Kant, we make something our maxim through an act of free choice. This is clear from his official definition of a maxim as "a subjective principle of action . . . [that] contains a practical rule *determined by reason* in accordance with the conditions of the subject (often his ignorance or again his inclinations). . . ."[4] Moreover, whatever considerations motivate us to adopt this or that maxim, these considerations are, qua motivating reasons that figure in a rationalizing explanation of maxim adoption, something that the agent chooses. This is made clear in Kant's so-called 'incorporation thesis.' This thesis is at the heart of Kant's theory of freedom of the will, and functions to ensure that rationalizing explanations are ultimately not just mechanistic causal explanations of actions. Kant states the thesis as follows:

> [A]n incentive can determine the will *(Willkur)* 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).[5]

---

[3] This is rough and taken at face value is not plausible. No doubt one's happiness depends on the satisfaction of so-called 'informed desires' — those desires that one would have were one to engage in appropriate reflection on her current desires in light of relevant information.

[4] Grundlegung zur Metaphysik der Sitten IV: 421, 88n, my emphasis. Also in the Grundlegung at 427, 94 Kant points out that desires and inclinations (settled desires) arise from feelings of pleasure and displeasure, and from these "with the cooperation of reason, there arise maxims." In the Kritik der praktischen der Vernunft V: 79, 82, Kant claims that an *interest* (which an agent takes up as a result of reflecting on her desires) is the basis, or provides the reason, an agent has for adopting a maxim.

Incentives, in Kant's terminology are what I have been calling motivating reasons for action. Kant's point here is that in order for any of the sorts of factors, including desires and emotions as well as beliefs the agent has about her situation and action — factors that may figure in a rationalizing explanation of the agent's actions — to count for the agent as a motivating reason, the agent must 'make,' as Kant says in a number of places, that factor a motivating reason. This claim holds not only for those incentives stemming from our sensuous natures, but holds as well for moral incentives that have a purely rational basis. That even reason-based considerations, in order to be motivating reasons, must be incorporated by the agent is made clear when, in characterizing the predisposition to personality as a capacity to be motivated by respect for the moral law (via moral feeling), Kant remarks that such an incentive becomes a motivating reason "only when the free will *(Willkur)* incorporates such moral feeling into its maxim. . . ."[6] Thus, for Kant, desires (and incentives generally) are not, independently of an agent's choices, motivating reasons for action, though, as I shall explain a bit later, Kant does think that independently of an agent's exerting her free will, certain considerations or incentives are justifying reasons for action.

4. The Structure of Rationalizing Explanations

To provide a rationalizing explanation of an action involves citing those motivating reasons of the agent that were effective in bringing about the action in question. Kant's view about the structure of rationalizing explanations is a foundational one, where the terminus or foundation of any complete and full rationalizing explanation of an action is what Kant calls one's disposition *(Gesinnung)*. The pattern of explanation, then, for any action would go as follows. The immediate link explaining why an agent performed some 'outer' action is the agent's intention or maxim of action. Maxims of actions themselves are adopted for reasons, and normally, the immediate link in explaining why the agent adopted some particular maxim of action will involve another maxim(s) the agent has adopted plus certain of the agent's beliefs. In the simplest kind of case, then, what explains (let us suppose) why Alex adopted the maxim of faithfully visiting his aged grandmother is the fact that he intends to impress his fiancee, Gretchen with his (apparent) concern for family, and believes that by visiting his aged grandmother he will indeed impress her. Alex's intention to impress Gretchen represents a maxim of ends, his intention to visit his grandmother represents a maxim of action, and rationalizing explanations typically explain a person's actions (and intentions producing those actions) in terms of goals, ends or projects that the agent has set for himself. Of course, for maxims of ends like the one in question, we can inquire about the agent's motivating reasons for adopting

---

[5] Die Religion innerhalb der Grenzen der blossen Vernunft VI: 23-24, 19.

[6] Ibid., VI: 27, 32.

it, and typically the explanation will be in terms of more general ends the agent has adopted. The resulting picture is a hierarchy of maxims where an explanation of specific maxims of action (where the action is within one's immediate voluntary control) appeals to other maxims organized in a series where the terminus of the series is some very general maxim. If the maxim whose adoption is to be explained involves, in the order of explanation, appeal to the most general maxim pertaining to the aim of one's own happiness, then the maxim (and action flowing from it) stem from desire-based reasons for action. Maxims whose explanation ultimately stem from a maxim to comply with the demands of morality are maxims whose ultimate explanation is in terms of reason-based reasons for action.

However, since there are two types of motivating reasons for action, and since in some cases at least one might take there to be two sorts of reasons that bear on explaining some action, one can ask which of these types of reason was actually efficacious and why. For instance, if I have a desire-based reason for doing A, but I also have (and recognize) a reason-based reason for refraining from A, then if I go ahead and do A, then we can ask for a contrastive rationalizing explanation of my action: we can ask why I did A rather than refrain from doing A. In this sort of context, merely citing a series of motivating reasons stemming ultimately from a desire for happiness is not enough for the sort of explanation sought. We want to know, in addition, why the agent, given competing motivating reasons for action, chose to do A rather than refrain from doing A. Explanations that answer this contrastive question may advert to all sorts of phenomena, including such things as lack of willpower, ignorance, and so forth. But here, the terminus of explanation is what Kant refers to as one's disposition, "the ultimate subjective ground of the exercise . . . of man's freedom in general."[7]

5. Disposition (Gesinnung) and the Supreme Maxim

In addition to the foundational status of one's disposition, there are four other important features characteristic of one's disposition. First, since one's disposition is something that bears directly on the morality of one's character, it can be either good or evil. Second, the moral quality of one's disposition is only revealed in a series of choices in which one is faced with moral obligations. If, on some lone occasion, one does one's duty solely for duty's sake, one does not thereby have either a good or an evil disposition. Thus, were one capable of a God's-eye view of people and their motivation, attribution of a good or an evil disposition would require viewing the agent's choices diachronically. Third, since one's disposition concerns the two sources of motivating reasons for action and whether, in cases involving an agent having a moral duty, one does or does not act on the basis of purely moral incentives, there are only two basic motivational orientations associated with the notion of disposition. If one views one's choices

[7] Ibid., VI: 21, 16.

diachronically, then in all cases where one has a duty, either (i) one's sole motivating reason for action is a moral one, or (ii) it is not. This is Kant's so-called character rigorism: with regard to one's disposition (and hence the deepest aspect of one's character) one's disposition cannot fail to have a moral quality, and cannot, at a time, be both good and evil. Fourth, Kant claims that one is responsible for one's disposition; it can be imputed to the agent. This claim is an implication of Kant's theory of freedom: if our adoption of specific maxims can be imputed to us, and if the ultimate subjective ground, as Kant calls it, provides the terminus in any full rationalizing explanation, then this ground must itself be imputed to the agent as something under her voluntary control. Thus, according to Kant, our disposition can be imputed to us as the result of the exercise of one's free choice. Furthermore, since such exercise results in maxim adoption, Kant characterizes an agent's disposition as a maxim, and given its foundational status in the chain of motivating reasons for action, he calls it a 'supreme maxim.'[8]

To summarize: Maxims are intentions adopted by agents for reasons. These reasons — motivating reasons — become reasons as a result of an agent's giving them motivational influence (the Incorporation Thesis). There are two species of motivating reason, viz., desire-based and reason-based, and reasons of both sort figure in rationalizing explanations of agent's maxim adoption and resulting action. A full and complete explanation of any action will advert to what Kant calls the supreme maxim, representing one's most fundamental motivating orientation *(Gesinnung) vis a vis* the two species of reasons for action. The morality of one's character depends on one's motivational orientation in a manner to be elaborated below.

II. Error in Practical Reason

Talk about an agent's acting rationally on some occasion is ambiguous, an ambiguity reflected in talk about an agent's reasons for action. If we take the agent's perspective, we can ask whether, from the agent's subjective perspective, some action she performed on some occasion was rational. Here, our answer to this question will focus on what reasons the agent took to be good or adequate ones, whether or not they really are. On the other hand, we might take an external perspective appealing to objective canons and principles of rational conduct in

[8] The apparent tension between construing one's disposition as both the result of an act of free choice (for which, it seems, one must have a reason) and a terminus in a series of motivating reasons is the source of interesting puzzles for Kant's view. Kant recognized the tension and claimed that the choice of one's disposition must be represented as a timeless noumenal act whose rationale must remain inscrutable to us. Here, I pass over this difficult issue, though, following *Allison* (op. cit. f. 2), pp. 135-140 I am inclined to interpret this doctrine of a timeless noumenal choice in a minimalist way as committing Kant only to a claim about the limits of rationalizing explanations.

assessing the rationality of an agent's behavior. In this section, my remarks will concern objective assessments of rational action. Thus, for an agent to act rationally on some occasion (where the standards or principles of rationality being employed are what I am calling objective principles) involves, at a minimum, her having a good or justifying reason for performing that action on that occasion, moreover a justifying reason that is better or weightier than any other reasons an agent may have that favor doing something else. Furthermore, if the act is to be rational, the agent must act in light of her justifying reasons, where acting in light of such reasons involves reasoning according to principles of practical reasoning — practical principles of the sort that Kant calls "objectively valid" — such reasoning figuring in a rationalizing explanation of the agent's action.

As I mentioned in my introductory remarks, for Kant, evil represents a kind of irrationality connected with practical reasoning and action. In this section, I want to focus on just what sort of irrationality is involved in evil doings.

Kant distinguishes two sorts of practical principles: *subjective practical principles,* or maxims, and *objective practical principles.* Objective practical principles are principles of practical reasoning — principles that govern the rational revision of one's maxims or intentions. As such, objective practical principles can be used to guide revision of one's own set of maxims, they can also be used to judge the rationality of an agent's maxim adoption. Because of their status as standards of practical rationality, these principles are, like principles of logic, objective or valid, as Kant would say, for assessing and guiding the maxims adopted by agents. Furthermore, objective principles reflect considerations in virtue of which one has justifying reasons to perform some action (and hence in virtue of which one has reason to adopt the corresponding intention or maxim to perform that action or adopt some end). Kant's formal principle of hypothetical imperatives — what we may call his *principle of heteronomy* — is best interpreted as a principle of practical reasoning counseling an agent to adopt those maxims of action that are necessary for achieving those ends or goals that she has reason to achieve. For example, if I have reason to loose weight and intend to do so, and if reducing my intake of carbohydrates is necessary for my losing weight, then the principle of heteronomy counsels me to either adopt the maxim of reducing carbohydrates or give up my end, on pain of irrationality.

Moral evil represents for Kant a kind of practical irrationality. To perform an action that falls short of full rationality is to be guilty of some error or mistake, an error that can often be traced to one's practical thinking. So, evil behavior must involve some sort of error in practical reasoning. I suggest the sort of error involved here concerns an agent violating an objective practical principle. Let me elaborate.

We can think of errors in practical reasoning as involving violations of practical principles. There are two sorts of possible error connected with practical thinking corresponding to two phases of deliberation. If practical thinking aims at revision

of one's intentions, we can distinguish the phase that precedes the actual revising of one's intentions from the second phase in which one actually engages in revision. In this first phase, there is a process of reflection in which one engages in some or all of the following: envision alternative courses of action, consider possible outcomes of each action, try to vividly imagine what the various possible outcomes would be like, reflect on those considerations that favor doing one act over another, weigh various considerations that favor one action or another in order to determine which considerations provide the best reasons for action, and so forth. I will call those rules that specify what sorts of things one should consider in this reflective phase, *rules of reflection.*[9]

In the second phase of deliberation, one actually revises one's intentions — adopts new maxims, perhaps revises or even rejects current maxims. I will call those rules or principles governing correct revision of one's maxims, *principles of revision.* Kant's principle of heteronomy, then, can be understood as a principle of revision prescribing how agents ought to revise their set of intentions given certain means-ends connections.

Rules of reflection, then, prescribe the various sorts of things one should consider in practical thinking; principles of revision put constraints on how one is to revise one's intentions. Error in practical reasoning can be understood as a violation of one or more of these rules and principles. Interestingly, it would seem that the Kant's principle of autonomy (the Categorical Imperative) functions both as a rule of reflection and as a principle governing correct revision of one's set of intentions. For example, in one place Kant says that the principle of autonomy can be expressed as claiming that "Maxims must be chosen as if they had to hold as universal laws of nature."[10] Moreover, according to Kant, we are to *will,* i. e., adopt and act on maxims, in accordance with the principle of autonomy.[11] Such passages support the idea that imperatives generally, and specific categorical imperatives in particular, are principles of revision, governing correct modification of intentions. However, other passages where Kant claims that the principle of autonomy functions at least implicitly in people's moral thinking as a "norm of judgment"[12] make clear that Kant also thinks of the principle of autonomy as a rule of reflection, governing correct judgment or belief about moral obligation.

---

[9] See *Gilbert Harman,* Change in View, Cambridge, MA: The MIT Press, 1986, ch. 1, who distinguishes between two phases of practical reasoning (whether theoretical or practical) and consequently between two sorts of errors of reasoning: errors of reflection and errors of revision. He notes that there are other sorts of mistakes that one might make while reasoning, like starting with false beliefs, but these are not, he claims, errors of reasoning.

[10] Grundlegung zur Metaphysik der Sitten IV: 326, 104.

[11] Ibid., IV: 416, 84.

[12] Ibid., IV: 404, 71; cf. 390, 57.

It should be clear that the deontic (or as Kant would say, the 'legal') status of actions depends crucially on whether or not the agent violates any of the rules and principles governing correct practical reasoning. Most obviously, if one fails to adopt (and act on) universalizable maxims, then one violates the principle of autonomy — violates a principle of revision. But also, one's action may fail to be rational (and hence is forbidden) if one violates those rules of reflection that would require one to take care in avoiding e. g., bias in formulating maxims for moral consideration. At least both sorts of error can be recognized within Kant's theory of practical reasoning, and, as we shall see below, failure to properly represent the rational weight of moral considerations represents a failure of rationality that is at the root of wickedness.

The purpose of this section, and the one preceding it, has been to set the stage for making sense, according to Kant's view of practical reasoning, of moral evil manifested in actions that are imputable (done for reasons) yet irrational. Let us then proceed to Kant's account of moral evil.

III. Error and Evil

Kant claims that on those occasions where we are morally required to do something, we ought to do what is morally required from the sole motive of duty. Failure to do so involves, then, a kind of error — an error of the sort that indicates, for Kant, moral evil. Moreover, failure to perform one's duty from the motive of duty can be manifested in more than one way. In book I of the *Religion,* Kant describes three types of moral evil that he labels frailty, impurity, and wickedness. These terms are applied both to actions and to a person's character. If, on some occasion, one fails, through weakness of will, to do what one recognizes one ought to do, one's action exhibits moral frailty. But a single instance of such frailty does not make one a morally weak or frail person, rather, only if such weakness is characteristic of a person's choices is it correct to say that one has a morally weak character. In what follows, the focus will be on prototypical cases of evil action of the sort Kant recognizes.[13]

[13] There are other alleged types of moral evil that are regularly featured in philosophical discussions, including various forms of moral negligence and amoralism. The case of the amoralist (someone who recognizes that she has a moral obligation to do A, but is not at all motivated to do A) is a particularly interesting one for Kant, since if one construes Kant as an ethical internalist (according to which, roughly, there is a conceptual connection between obligation and motivation), then one can't allow for the possibility of the amoralist. I think there are broad historical reasons for construing Kant as an internalist (see *Mark Timmons,* "Kant and the Possibility of Moral Motivation", The Southern Journal of Philosophy 23, pp. 377-98), but also there is some textual evidence. In Die Metaphysik der Sitten VI: 379, 185n, Kant writes: "Yet if man looks at himself objectively (under the aspect of *humanity* in his own person), as his pure practical reason determines him to do, he finds that *as a moral being* he is also holy enough to break the inner law *reluctantly;* for there is no man so depraved as not to feel an opposition to breaking it and an abhorrence in himself in the fact of which he has to constrain himself [to break the law]." I read this as claiming that necessarily all moral agents are such that recognition of the moral law is motivating (even though, as Kant goes on to remark in this same passage, other, non-moral motives may get the motivational upper hand on one's moral motivation). Obviously, there is a great deal more to be said about Kant's notion of moral evil than I can hope to cover here.

In order to understand moral evil, we need to answer, in connection with each of these types of evil, the following questions: (1) In what sense does behavior manifesting these failures represent a failure of rationality? (2) How can such irrational behavior be imputed to agents who engage in it? (3) What is the underlying nature of such evils; in virtue of what are frailty, impurity and wickedness evils? and (4) What sorts of psychological phenomena are at work in the production of such these forms of irrational behavior. Most of my discussion will focus on answers to the first three questions; I shall only make some passing remarks in response to the fourth question which obviously calls for a complex psychological story that I cannot delve into here.

1. Moral Weakness

Kant describes this kind of character flaw and the choices it manifests as follows:

> [T]he frailty *(fragilitas)* of human heart is expressed even in the complaint of an Apostle, "What I would, that I do not!" In other words, I adopt the good (the law) into the maxim of my will, but this good, which objectively, in its ideal conception *(in thesi)*, is an irresistible incentive, is subjectively *(in hypothesi)*, when the maxim is to be followed, the weaker (in comparison with inclination).

The sort of frailty or moral weakness that Kant has in mind is a species of the general phenomenon of weakness of will. Normally, a person manifests moral weakness whenever that person: (i) judges that some action is morally required in some situation; (ii) takes this fact to provide her with an overriding reason for action (i. e., ranks the reason provided by this moral requirement above other, competing reasons); (iii) she consequently incorporates the moral incentive into her maxim, as Kant would say, i. e., she makes the moral incentive a motivating reason for action; (iv) were there no competing reasons, the moral reasons would be sufficient to motivate the agent to perform the required action in the situation; but (v) she fails to act according to what she judges she has most reason to do; yet (vi) her failure can be imputed to her; and (vii) as a result she experiences feelings of guilt and remorse.[14]

[14] I am not proposing a necessary and sufficient conditions analysis of the concept of moral weakness of will, since, for one thing, I don't think (vii) is a necessary condition of moral weakness, but rather is only associated with typical cases and functions epistemically to indicate that the agent knowingly engaged in moral wrongdoing. So, I intend (i)-(vii) as a description (in partly Kantian terms) of a prototypical case of this sort of failing.

In what sense, then, does such behavior involve an error in practical reasoning in virtue of which it is irrational behavior? And in what sense can such behavior be imputed to the agent?

From what was said in the previous section, the answer to the first question is straightforward. If we view the principle of autonomy as a principle for rationally revising one's maxims, then cases of moral weakness, in which one knowingly fails to revise one's maxims according to this principle result in irrationality. That is, the principle of autonomy requires that agents adopt only universalizable maxims on pain of irrationality, and so one way in which the morally weak person's action is irrational is simply that it is morally forbidden ('illegal', as Kant would say).

But there is also another sense in which the morally weak person's action is irrational. Another sort of failure of practical rationality results from what we might call a lack of correspondence between those reasons for action that are the best reasons (in the sense of representing considerations having the most rational weight) and a person's strongest motivating reasons. In talking about justifying reasons for action, we may rank them according to the rational weight they possess. Some justifying reasons for action are weightier than others, and we can talk about some reasons *outweighing* (in terms of rational strength) other reasons. In cases of conflicting reasons for action, i. e., where, on some occasion, one reason R′ favors doing some action A, and some other reason R″ favors doing some other action B, and R′ is a better reason than R″, let us say that R′ *overrides* R″. Talk of some reasons outweighing and overriding other reasons is to be taken, then, as concerning the rational weight or *authority* of those reasons. When it comes to motivating reasons for action, we can talk about the *strength* of those reasons, where talk of strength refers to how strongly one is motivated by some consideration. Most obviously, desires vary in degrees of felt strength. In cases where one motivating reason is stronger than some other motivating reason, let us say that the first has (for the agent, on that occasion) *motivational dominance* over the second.

The lack of correspondence I have in mind, then, in connection with weakness of will concerns the relative weight of justifying reasons and their failing to match the relative strength of motivating reasons. That is, we might propose the following Principle of Motivational Correspondence:

PMC The strength of one's motivating reasons for action on some occasion ought (rationally) to correspond to the weight of one's justifying reasons for action on that occasion.

This principle can be used to explain why any case of weakness of will involves a kind of irrationality; appeal to the requirements of the principle of autonomy, together with PMC, can be used to explain why some particular case counts as a case of *moral* weakness of will. In cases of the latter sort, then, the principle

of autonomy provides justifying reasons that outweigh and override any other reasons for action, and insofar as one's choice (and consequent action) is based on non-moral motivating reasons, one's choice (and action) is irrational. As Kant says, in cases of moral frailty the fact that one has a moral reason to act in a certain way, "objectively, in its ideal conception," represents an overriding reason for action, yet it is "the weaker in comparison with inclination."

What about imputing such behavior to the morally weak agent? To explain such behavior as something the agent did, and hence as something freely done by the agent, we must be able to explain the action from the agent's perspective, i. e., by appealing to those features of the agent's psychological makeup that would serve as a rationalizing explanation of the behavior. Here, of course, Kant's Incorporation Thesis is important. Recall that according to that thesis, some motivating consideration can become a reason for action only through an agent's making it so. So, it would seem that immoral behavior characteristic of the morally weak agent can be explained in terms of the agent's reasons for action: she does have reasons for acting as she does — considerations bearing on action whose relevance depends on inclination — which she intentionally takes to be a reason for action (and thus, as Kant would say, "incorporates into her maxim"), and which thus provides a rationalizing explanation of her action. Her action (or omission) is thus imputable.

But the problem with this sort of rationalizing explanation is that it does not adequately account for the phenomenon in question. We are still left with a puzzle regarding imputation. Although the agent (in the typical case) does have a reason for omitting to do what is morally required, what needs really needs explaining here is *why the agent knowingly performed the worse act and not the one backed by the best reasons.*

To satisfy our constraint on accounting for moral weakness of will, I suggest that we must modify slightly our interpretation of the Incorporation Thesis. My suggestion is that we should allow for two sorts of incorporation, at least when it comes to non-moral motivation. One sort is where inclinations are taken by the agent to be reasons for action. But there is also the case where the considerations bearing on action prompted by inclination are not taken by the agent to be reasons for action, but nevertheless those considerations are allowed by the agent to have motivational push. Notice, that quite apart from the phenomenon of moral weakness, we have to be able to accommodate those cases where the agent acts out of motivational considerations that are not so-called 'endorsed' motivating reasons for the agent. (Recall my examples of the would be non-smoker and the sadist.) I suggest that weakness of will in general, and moral weakness of will, in particular, where what is to be explained (and imputed) is why the agent did what she recognized to be the worse course of action rather than the better, can be explained in terms of the second kind of incorporation. That is, in such cases, there are no considerations that the agent takes to be reasons for doing

the worse instead of the better act, however, she does allow non-moral considerations to have sufficient motivating strength, and in so doing, her resulting action can be imputed to her. Thus, by broadening the Incorporation Thesis, Kant can account for cases of moral weakness as irrational, but explicable behavior.[15]

What then is it about the morally weak agent in virtue of which she is evil? The answer for Kant lies in the sort of disposition of the agent that represents, in the series of rationalizing explanations, the terminus of such explanation. As I explained above, one's disposition represents one's most basic orientation of choice in relation to the two sources of reasons for action. Kant's conception of a morally good finite rational agent is one whose basic orientation is such that whenever one is morally required to perform some action (or omission), the moral incentive is the sole and sufficient motivating reason for action. Given Kant's character rigorism, if one is not morally good, then one is morally evil — has an evil disposition. Thus, the morally weak person has an evil disposition. Moreover, this sort of evil — evil at the level of one's disposition — involves a violation of the obligation to "act in conformity with duty *from* duty."[16] So although actions that manifest moral weakness are irrational for reasons mentioned above (they flow from non-universalizable maxims of action and violate PMC) they also involve a violation of the principle of autonomy at the level of one's disposition. It is this particular violation of the principle of autonomy, then, that accounts for the irrationality involved in the evil of moral weakness.

[15] What sorts of psychological mechanisms are at work that would explain why, in cases of moral weakness, the agent allowed certain desires and perhaps urges to have motivational dominance? One might appeal, as some have, to the phenomenon of self-deception to explain why the agent engages in such behavior. The idea would be that although the agent "knows in her heart" that the action she performs is morally wrong and that this fact about it provides an overriding reason to refrain from performing it, she nevertheless, through some process of self-deception, proceeds on the belief that the action in question is supported by the weightiest reasons and that, consequently, her action is not wrong. Although self-deception may be one mechanism that explains certain sorts of evil behavior, if my characterization of moral weakness is correct, self-deception can't be at the root of morally weak behavior. A morally weak person knowingly performs the worse act and consequently (in normal cases at least) has feelings of guilt. Henry Allison claims that in order to make sense of Kant's degrees of radical evil one must assume that self-deception is at the root of all three degrees and an essential ingredient of radical evil. In the case of moral weakness he claims that one self-deceptively depicts "what is in reality a free evaluation on one's parts as a 'weakness' for which one is not responsible." *Allison* (op. cit. f. 2), p. 159. But I find it hard to see why Allison thinks this. The analysis I provide of radical evil does not require that self-deception necessarily be involved either in connection with the agent's beliefs about the deontic status of the action she performs or about her responsibility for it. One plausible suggestion about the psychological mechanisms underlying self-deception offered by Ronald Milo appeals to lack of willpower on the agent's part. See *Ronald D. Milo,* Immorality, Princeton: Princeton University Press, 1984. Talk of willpower, for Milo, refers not to some mysterious, hidden capacity that we have for overcoming temptation, but refers rather to a battery of mostly acquired skills (such as reminding ourselves of our reasons for avoiding certain behavior and how we will feel afterwards if we cave in to temptation, and so forth) that we use to manage the influence of our desires and emotions one our choices. However, as I said at the outset of this section, theorizing about the psychological mechanisms at work in cases of weakness of will is not our foremost concern here, and so I leave it open just what sorts of mechanisms (compatible with Kant's view) are operative.

[16] Die Metaphysik der Sitten VI: 391, 194.

2. *Moral Impurity*

Kant describes this form of irrational behavior as follows:

> [T]he impurity *(impuritas, improbitas)* of the human heart consists in this, that although the maxim is indeed good in respect of its object (the intended observance of the law) and perhaps even strong enough for practice, it is yet not purely moral; that is, it has not, as it should have, adopted the law *alone* as its *all-sufficient* incentive: instead, it usually (perhaps, every time) stands in need of other incentives beyond this, in determining the will to do what duty demands; in other words, actions called for by duty are done not purely for duty's sake.[17]

Whereas cases of moral weakness involve immoral or evil wrongdoing, cases of moral impurity do not involve wrongdoing at least as regards the action one performs — the agent acts on a universalizable maxim and so her action is not morally forbidden. However, cases of moral impurity are quite similar to cases of moral weakness in that they manifest the same sort of weakness explained above.[18] To see this, we first need to notice that from Kant's description of moral impurity, he has in mind cases where a person does have a moral motivating reason for doing her duty, but she nevertheless, as Kant puts it, "stands in need of other incentives beyond this" in order to do her duty. In other words, for the morally impure agent, although she has made the moral law a motivating reason, and perhaps, at times, does do what is morally required out of her sense of duty, nevertheless, often enough for her, the moral incentive is not sufficient to move her to comply with duty.[19]

Now Kant claims that the moral law would always be a sufficient motivating reason for action were there not some other, non-moral motivating reason that the agent allows to have motivational dominance over moral motivating reasons. He writes: "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

[17] Die Religion innerhalb der Grenzen der blossen Vernunft VI: 29-30, 25.

[18] In the Grundlegung zur Metaphysik der Sitten IV: 406, 74, Kant mentions frailty and impurity as modes of human wickedness.

[19] For brevity's sake, I am ignoring cases of motivational overdetermination in which an agent does her duty, though she had both moral and non-moral incentives for doing so and incentives of both sort were operative (would figure in a rationalizing explanation of her action). One might claim that in such cases, the agent manifests the sort of moral impurity Kant is talking about here, since the motive of duty was not the sole motive for action.

of his will *(Willkur)*."[20] So, in cases of moral impurity, although the agent has made the moral law a motivating reason for action, and indeed, it is even sufficient in cases where no other, competing non-moral motivating reasons are operative, the moral incentive fails to be sufficient because of competing reasons that favor not doing what is morally required. If the agent does in fact perform the morally required action, it will only be because, in addition to those non-moral reasons that favor not doing the required action, the agent also has non-moral reasons that favor performing the dutiful action, and these have motivational dominance over the competing reasons. So, for example, my own laziness and the attraction of laying around watching television may be strong enough on some occasion that unless I take myself to have some very good prudential reason for doing A (which just happens to be my duty), I will not do A. In cases where I do have such a prudential reason, I end up doing my duty, but not for moral reasons.

Given these remarks, we can characterize the prototypical case of moral impurity as involving an agent who: (i) recognizes that he has a moral requirement to do A in circumstances C; (ii) takes the fact that doing A in C is morally required to provide an overriding reason to act accordingly; (iii) adopts the maxim of doing A in C; however (iv) the agent also has non-moral reasons that are sufficient to move him to do A in C; and (v) these non-moral considerations (rather than moral reasons) figure in a rationalizing explanation of why he adopted the maxim in question.[21]

If this form of moral evil is irrational, what sorts of errors in practical reasoning are involved? Given my description of this sort of case, the agent need not be violating any rules of reflection: he recognizes that he is morally obligated and he takes that fact to provide an overriding reason for action. The practical error, then, occurs in connection with the principle of autonomy — a violation of a principle of revision. However, unlike the case of moral weakness, an agent whose will manifests impurity on some occasion does not violate the principle of autonomy as it relates to the deontic status of the person's action, but rather violates that principle as it applies to one's underlying subjective principle of motivation, i. e., one's supreme maxim. That is, one violates the obligation to act from the motive of duty. Agents who knowingly fulfill their moral obligations, but who need (at least on occasion) additional motivational spark from non-moral considerations in order to do so, fail to have a certain orientation of the will, which, for Kant, means that they fail to have the right supreme maxim.

[20] Die Religion innerhalb der Grenzen der blossen Vernunft VI: 36, 31.

[21] To this characterization we might add that although the agent will not experience feelings of guilt associated with his recognition of the deed as morally required (he did, after, comply with the demands of morality), nevertheless, the agent may experience feelings of guilt associated with the manner in which he complied with duty. The reflective agent will realize that the rational force of moral requirements fails to be matched by his actual motivation. Of course, the agent may not be so reflective, or may engage in a form of self-deception in which he hides from himself his real motivation. However, not all cases of moral impurity need involve such self-deception.

Moreover, like cases of weakness of will, cases of moral impurity violate the principle of motivational correspondence. In those cases where the agent's moral incentives are not sufficient to move her to perform the dutiful act and she needs a non-moral push to do so, the strength of an agent's moral motivating reasons fail to correspond to their rational status as overriding reasons.

So, if my characterization of moral impurity is correct, then this sort of evil involves a kind of weakness on the agent's part: like the prototypical morally weak person, the morally impure person's moral motivation is too weak in the face of other, competing concerns to move her to action, though, luckily, she does have sufficient non-moral reason to take up the motivational slack. The essential difference, then, between morally weak behavior and morally impure behavior is one of moral luck — something that impure behavior manifests that morally weak behavior does not.

Making sense of how impure action can be imputed is unproblematic (bracketing, for course, questions about imputing one's disposition). Unlike the case of moral weakness, the agent does perform the act that he judges he has most reason to do, even though he does so by luck: one's non-moral reasons for action just happen to move one to perform one's duty. However, there is a moral fault involved here, since the agent fails to comply with a second-order duty to act from duty. This fault concerns one's basic motivational orientation and explains why moral impurity is a species of moral evil. As we saw above in connection with moral weakness, since the maxim of acting from the sole and sufficient motive of duty (in situations where one has moral obligations) fails to be adopted by the agent as his supreme maxim, the agent fails to have a morally good character or disposition, and is thus (given Kant's character rigorism) necessarily morally evil.

3. Wickedness

Kant describes wickedness as involving an inversion of the proper ordering of the reasons for action:

> [T]he wickedness *(vitiositas, pravitas)* or, if you like, the *corruption (corruptio)* of the human heart is the propensity of the will to maxims which neglect the incentive springing from the moral law in favor of other which are not moral. It may also be called *perversitas)* of the human heart, for it reverses the ethical order [of priority] among the incentives of a *free* will.[22]

For the morally wicked person, non-moral reasons for action enjoy motivational dominance over moral reasons for action (which may also be true of the morally weak person) but such motivational dominance is *principled.* That is, not only does the wicked person fail (as in the cases of weakness and impurity) to have

[22] Die Religion innerhalb der Grenzen der blossen Vernunft VI: 30, 25.

the right supreme maxim, but she has in effect deliberately adopted a supreme maxim that gives priority to non-moral reasons. Whereas cases of moral weakness always involve failing to perform an obligatory action, and cases of impurity are described as cases where the agent fulfills the 'letter' of the moral law and so does the obligatory action, wickedness may or may not involve violations of one's moral obligations. The sort of person who fits Kant's characterization here is someone who: (i) recognizes moral requirements, (ii) not only fails to allow such requirements to have sufficient motivational force, since as Kant says, "in this case no attention whatever is paid to the motivating forces in the maxim,"[23] but also (iii) in so denying them their proper motivational role, the agent is committed in a principled way to pursuing non-moral ends, regardless of how they might conflict with moral ends. Making pursuit of non-moral ends a matter of principle is a matter of having adopted a very general maxim that, as Kant would say, deliberately reverses the proper order of the two basic sorts of reasons for action.

Kant makes two other claims about wickedness. First, he thinks that wickedness is somehow morally worse than the other two forms. For instance, he claims that as a result of moral evil (represented as something freely chosen), we have a kind of innate guilt of two sorts: "this guilt may be judged in its first two stages (those of frailty and impurity) to be unintentional guilt *(culpa)*, but in the third to be deliberate guilt *(dolus)* and to display in its character a certain insidiousness of the human heart *(dolus malus)*."[24] Second, he claims that the wicked person engages in self-deception "in regard to its own good and evil dispositions, and, if only its conduct has not evil consequences . . . does not trouble itself about its disposition, but rather considers it justified before the law."[25]

How can this 'high octane' form of moral evil be represented as both irrational and imputable? And how can we explain the insidious nature of wickedness that sets it apart from the other two forms? Moreover, how is self-deception involved in prototypically wicked behavior and character?

In order to answer these questions, we should note that from what Kant says about wickedness, the wicked person is apparently someone for whom there is a principled failure of moral reasons to have motivational dominance because that person fails to properly represent the rational authority of moral considerations. That is, the wicked person Kant seems to have in mind fails to rank the reasons for action associated with moral requirements over non-moral reasons.[26]

[23] Ibid., VI: 38, 33.

[24] Ibid., VI: 38, 33.

[25] Ibid., VI: 38, 33.

[26] Actually, in Kant there seem to be at least two cases of moral wickedness to sort out. In addition to the case just described where a perverse value judgment is at the bottom of one's evil disposition, there is the case of moral negligence, where the agent fails to recognize her duty to discharge moral obligations from the motive of duty. The morally negligent person, as I am calling her, may well properly rank moral requirements (requirements featured in common duties to ourselves and others) above non-moral ones, and she may even be disposed to give motivational dominance to moral reasons (perhaps she is moved by a strong sense of sympathy toward others). Her problem is that she fails (perhaps through self-deception) to recognize what Kant claims is the most basic duty regarding one's motivation — to make the moral law the sole and sufficient motive of dutiful action. This case of moral negligence falls under my generic description of wickedness, and some of Kant's remarks in the Religion VI: 38, 33 suggest that this sort of failure is a form of wickedness, though it obviously differs from the case I'm describing in the text.

If this is right, then there is a double sense in which the wicked person is irrational. First, and most obviously, one violates the principle of autonomy as a principle of rational maxim revision in having adopted a maxim that gives practical priority to non-moral reasons for action. Again, this character defect can be imputed to agents since it is represented as something one chooses — a maxim one adopts. Moreover, as we have seen in connection with weakness and impurity, this sort of irrationality is the basis of evil in a person's character.

But secondly, in misrepresenting the proper authority of reasons stemming from moral requirements, one is violating a principle of reflection. Recall that principles of reflection govern the rationality of the first phase of practical thinking in which one does such things as consider alternative actions, balance and weigh reasons for various courses of action, and so forth. Moral requirements are overriding reasons for action, and failure to accord them that status in one's practical thinking is in violation of a rule of reflection — a rule in effect requiring that one properly represent to oneself the various weights attaching to those considerations bearing on thought and action. Indeed, this kind of irrationality associated with wickedness is what seems to set this form of evil apart from the other two. That is, whereas in cases of weakness and impurity, the agent at least correctly represents to herself the importance or rational weight attaching to moral considerations *vis a vis* other, non-moral considerations, the wicked agent fails to properly represent to herself the importance of moral considerations. In short, the wicked agent's choices are based on a perverted value judgment.

One likely explanation for this perverted value judgment is self-deception. Since, for Kant, the dictates of morality are experienced as categorical requirements, and indeed are naturally experienced by all agents in this way, it is only through something like self-deception, where one somehow gets oneself to believe that moral requirements do not have overriding authority, that one can end up misrepresenting the true authority of these requirements. Importantly, there are places, particularly in the *Grundlegung*, where Kant mentions the rational authority of moral requirements and the self-deceptive ploy of "juggling with conscience or with other claims as to what is to be called right, or in trying to determine

honestly for its own instruction the value of various actions. . . ."[27] A paragraph later, Kant writes:

> Man feels in himself a powerful counterweight to all the commands of duty presented to him by reasons as so worthy of esteem — the counterweight of his needs and inclinations, whose total satisfaction he grasps under the name of 'happiness'. But, reason, without promising anything to inclination, enjoins its commands relentlessly, and therefore, so to speak, with disregard and neglect of these turbulent and seemingly equitable claims (which refuse to be suppressed by any command). From this there arises a *natural dialectic* — that is, a disposition to quibble with these strict laws of duty, to throw doubt on their validity or at least on their purity and strictness. . . .[28]

Calling into question the authoritativeness (validity) of moral requirements through a process of 'quibbling' with that authority presumably leads, if one is a good enough quibbler, to the sort of perverse value judgment that I claim is at the root of the Kantian conception of wickedness.[29]

We can sum up this discussion of the three degrees of moral evil by considering how these character traits — frailty, impurity, and wickedness — involve a lack of moral commitment. The morally frail or weak person is someone whose motivating reasons for action are not in accord with her judgments about the authority of moral requirements: consideration of one's moral requirements which one recognizes provide the best reason for action, fails to have motivational dominance over competing, non-moral reasons. Impurity involves a different sort of lack of commitment: one's commitment to perform one's duty does not stem exclusively from moral considerations; one allows non-moral reasons to function as motivating reasons for complying with the demands of duty. Finally, wickedness involves a principled lack of motivational dominance: one not only fails to give motivational dominance to moral requirements, but this failure is based on a value judgment, viz., the judgment that moral requirements are less important than non-moral ones. The fact that this sort of evil involves a perverse value judgment at its root explains why being wicked is to be in a morally worse state than being either weak or impure. In each case, however, there is a failure of moral requirements to have motivational dominance or, put another way, in each such case, one fails to have as one's supreme maxim the maxim of doing one's duty for the sake of duty alone. This failure is what constitutes moral evil and is that in virtue of which each of the three degrees of evil are evil.

[27] Grundlegung zur Metaphysik der Sitten IV: 404, 72.

[28] Ibid., IV: 405, 73.

[29] In the Grundlegung at 424, 92, Kant claims that this natural dialectic (mentioned in the passage just quoted) often has the result that we "permit ourselves a few exceptions which are, *as we pretend,* inconsiderable and apparently forced on us." (My emphasis.)

IV. The Nature of Good and Evil Disposition

Given the discussion in the previous section, we can define the Kantian notions of good and evil disposition *(Gesinnung)*. Quite simply, to have a good disposition involves having adopted, as one's supreme maxim, the maxim of doing one's duty for the sake of duty (on the relevant occasions) and consequently having a standing commitment to moral concerns that outweighs other, competing non-moral concerns. Moreover, according to Kant, "Virtue . . . [is] the firmly grounded disposition strictly to fulfil our duty . . ."[30] and so this sort of standing moral commitment is the defining characteristic of the morally virtuous person. And, of course, since the possession of one's disposition (good or evil) is imputed to the agent, the notions of a good disposition and good will (as that notion is featured in chapter I of the *Grundlegung)* are the same.

Given Kant's character rigorism, the essence of an evil disposition then, involves a failure to have adopted the moral supreme maxim. A person with an evil disposition lacks a virtuous character. Kant claims that there are two types of non-virtuous character: those persons who merely lack virtue and those whose failure is principled and hence morally vicious. Persons merely lacking in moral virtue exhibit a kind of weakness which, Kant says "is not so much vice *(vitium)* as rather mere *want of virtue,* lack of moral strength *(defectus moralis)* . . . It is when an intentional transgression has become a principle that it is properly called a vice *(vituim)*."[31] Thus, persons merely lacking in moral virtue exhibit moral weakness and / or moral impurity; morally vicious persons are wicked.

Though there is some controversy about how to understand Kant's notion of radical evil (see section VI), this notion would seem to be equivalent to the notions of evil disposition and lack of virtue just described. In the *Religion,* book I, Kant describes radical evil as a *propensity to adopt evil maxims.* This propensity, Kant claims, "must in the end be sought in the will *(Willkur)* which is free, and therefore be imputed, . . . [and so] is morally evil."[32] Hence, it must be a maxim, and since it is the "ultimate ground of the adoption or the observance of our maxims,"[33] this propensity is one's supreme maxim. Kant describes the nature of this propensity as follows:

> 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 the *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 in that he reverses the moral order of the incentives when he adopts them into his maxim. He adopts, indeed, the moral law

[30] Die Religion innerhalb der Grenzen der blossen Vernunft VI: 23, 19n.

[31] Die Metaphysik der Sitten VI: 390, 194.

[32] Die Religion innerhalb der Grenzen der blossen Vernunft VI: 31, 26.

[33] Ibid., VI: 32, 27.

> 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 its supreme condition, he makes the incentive of self-love and its inclinations the condition of obedience to the law; whereas on the contrary, the latter, ought to have been adopted into the universal maxim of the will *(Willkur)* as the sole incentive.[34]

Thus, if my analysis is correct, the concepts of a good disposition, a good will, and a virtuous character are equivalent, as are the concepts of an evil disposition, an evil will, a character that lacks moral worth, and one who is possessed of a radically evil will.

In a few places, Kant claims that having an evil disposition "may coexist with a will which in general is good,"[35] and in the above quote he allows that "even the best" man may still be evil. This claim may seem to conflict with equating a good will with good disposition, given Kant's character rigorism. But in the *Grundlegung*, the good will that has unconditioned, absolute worth is a will that has a firmly fixed disposition to do duty for duty's sake, what he calls an 'absolutely' good will. And again, in the *Religion* Kant defines a good person as one who has "the law as it sole and sufficient incentive . . . *always*.[36] A will that is *in general good*, is not absolutely good precisely because the agent allows occasional "moral holidays."

V. Radical Evil and Moral Worth

Moral worth is something possessed both by individuals (qua moral agents) and their actions. The moral worth of the individual is determined by that individual's supreme maxim. So, for instance, in the *Religion*, Kant considers the person whose actions may conform to the letter of the moral law, though they are not done from duty, and writes: "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."[37]

The more interesting (and controversial) question about moral worth concerns actions. In the *Grundlegung*, the moral worth of an action is a matter of its having been performed from the sole motive of duty. Now Kant claims that an evil disposition "corrupts the ground of all maxims," which, given that this sort of disposition is the foundation or ground of all others (it represents the terminus in a rationalizing explanation) means that this disposition (the ground) is corrupt. But it may appear as if Kant is also denying that the actions flowing from such

[34] Ibid., VI: 36, 31-32.
[35] Ibid., VI: 37, 32; cf: 30, 25.
[36] Ibid., VI: 30, 25.
[37] Ibid., VI: 31, 26.

a corrupt or evil will can ever have any moral worth.[38] After all, if the moral worth of one's maxim (and the action that flows from it) depends on the kind of motivating reasons that stand behind the adoption of one's maxim, and thus, depends ultimately on one's disposition, then whatever moral quality attaches to one's disposition will be inherited, so to speak, by the maxim in question. Kant, after all, is both a character rigorist and an action rigorist.[39] Thus, it seems to follow that if one has an evil disposition, then all of the maxims one adopts (and actions that flow from it) will be evil, hence not good and so lacking in moral worth. So the argument might go.

However, if we characterize having an evil disposition as I have, then it becomes clear how, despite having such a disposition, we may still be capable of performing actions that have moral worth. Lack of a good disposition entails, given Kant's rigorism about character, possession of an evil one. Having an evil disposition is a matter of not having a proper orientation of one's reasons for action: one fails to have a fixed supreme maxim of giving motivational dominance to moral requirements. But such a failing does not entail that one has positively reversed, as a matter of fixed principle, the proper ordering of one's reasons for action; wickedness, that is, is not the only form an evil disposition can take. One might be morally committed to some degree (unlike the wicked person), and even, on occasion, perform some dutiful action because, and only because, it is one's duty, yet one still might, from time to time experience bouts of moral weakness. Of course, someone who was chronically weak of will and was never able to summon the moral effort required to overcome competing, non-moral reasons in the face of duty, would not perform morally worthy actions, since that person would chronically fail to do her duty. But one's moral weakness need not be so extreme.

Again, a person can, on occasion, fail to make moral reasons the sufficient motivating reasons for action, and so only conform her action to duty in case she finds non-moral reasons sufficient to motivate her to perform that action. Failure to have adopted a fixed supreme maxim that makes moral requirements sufficient for action is to have an evil disposition. But this does not mean that, on occasion, one might not summon the moral strength to conform to duty for duty's sake without the help of non-moral motivating reasons.

[38] In his, "Good and Evil Disposition," Kant-Studien 76, 1985, pp. 288-, *Daniel O'Connor* writes: "And the disposition (if evil) destroys the worth even of his good acts, showing them to be merely 'legal' not 'virtuous'," p. 293. O'Connor holds that Kant's character rigorism, together with fact that an agent's character is good or evil, entails that actions flowing from that person's character have the same moral quality (good or evil) as her character. *Emil Frankenheim* ("Kant and Radical Evil," University of Toronto Quarterly 23, 1954, pp. 339-353) on the other hand, writes: "Thus, the motive behind an individual action may be respect for duty; and it may yet flow from an over-all maxim which includes the deviation from duty on other occasions," p. 349. I side with Frankenheim.

[39] See Die Religion innerhalb der Grenzen der blossen Vernunft VI: 22, 18.

What is not possible, on Kant's view, is for someone to have a wicked disposition, and hence a principled devotion to non-moral aims, and yet on any occasion perform an action from the motive of duty. The actions of wicked individuals can never be morally worthy.

VI. The Universality Thesis

According to the Universality Thesis (UT), all human beings suffer from the sort of defect in character that Kant refers to as radical evil. As Kant says, "the propensity to evil in mankind is universal, or what comes to the same thing, . . . it is woven into human nature."[40] Hence this thesis is one of strict universality; a necessary truth, knowable a priori. Kant claims that the thesis is not true in virtue of the concept of humanity, that is, it is not, in Kant's mind, an analytic truth; so it must be synthetic. And, of course, this means that it is a synthetic a priori statement requiring a special 'deduction.' However, Kant's apparent defense of UT simply appeals to empirical evidence: "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."[41] Perhaps Kant thought that given his audience, he did not need to provide a deduction of the UT, and that a few examples would suffice.[42] If the UT is supposed to be synthetic a priori, we do not find an explicit deduction of it anywhere in Kant's writings. However, Allison attempts to provide the needed, missing argument on Kant's behalf. He writes:

> The key to this deduction is the impossibility of attributing a propensity to good to finite, sensuously affected agents, such as ourselves (either to the race as a whole or to particular individuals). This impossibility, together with rigorism, entails the necessity of attributing a universal propensity to evil to agents relevantly like ourselves. And since, as we shall see, the impossibility at issue is not logical, (the notion of a propensity to good is not self-contradictory for Kant), the conclusion has synthetic a priori status.[43]

As I reconstruct Allison's deduction, the main steps are these. First, Allison provides a definition of a propensity to good, and an argument for the claim that human beings are not capable of a such a propensity. Second, this argument serves to defend a crucial premise in Allison's master argument described in the

[40] Ibid., VI: 30, 25.

[41] Ibid., VI: 32-33, 28.

[42] Though two paragraphs after the one containing the quote just mentioned, Kant writes: "But even if the existence of this propensity to evil in human nature can be demonstrated by experiential proofs of the real opposition, in time, of man's will *(Willkur)* to the law, such proofs do not teach us the essential character of that propensity or the ground of this opposition." Ibid., VI: 35, 30-31.

[43] *Allison* (op. cit. fn. 2), p. 155.

above passage: from the impossibility of human beings possessing a propensity to good, together with Kant's rigoristic thesis about character, viz., that a person's *Gesinnung* is either good or evil, to the conclusion that human beings are necessarily evil. Crucial to Allison's deduction is his supporting argument for the claim that human beings are incapable of a propensity to good. The impossibility here concerns the fact that human beings are susceptible to moral requirements. It will be helpful to make this argument (and the master argument) explicit.

Allison characterizes the sort of goodness in question this way:

> [A] propensity to good would consist in a kind of spontaneous preference for the impersonal requirements of morality over one's own needs as a rational animal with a built-in desire for happiness. . . for such an agent, the moral incentive would, as a matter of course, always outweigh the incentive of self-love. Consequently, for an agent blessed with such a propensity, there would be no temptation to adopt maxims that run counter to the law and, therefore, no thought of the law as constraining. Within the Kantian framework, this means that the law would not take the form of an imperative and moral requirements would not be viewed as duties.[44]

From this passage, we can extract a definition of a propensity to good:

(G) An agent has a propensity to good =df: that agent has a spontaneous preference for moral requirements; i. e., lacks a susceptibility to temptation to be motivated by considerations of self-love.

So the crucial supporting argument for Allison's master argument can be spelled out as follows:

1. Necessarily, for any finite rational agent, such agents have duties.
2. A duty (by definition) is an action to which we are obligated and thus (by definition) one that we are practically necessitated (constrained) to perform.

Therefore (from 1 and 2):

3. Necessarily, for any finite rational agent, such agents are practically necessitated to perform certain actions.
4. An agent is practically necessitated to perform an action only if she is susceptible to temptation (to act from non-moral motives). (This is a conceptual claim about the very notion of practical necessitation.)

Therefore (from 3 and 4):

5. Necessarily, for any finite agent, she is susceptible to temptation to transgress duty out of self-love (from non-moral motives).

Premise (1) is a synthetic a priori proposition — in fact it is the proposition that Kant attempts to demonstrate in *Groundwork* III.[45] Premises (2) and (4) are

[44] Ibid., 155.

[45] For a detailed discussion of this claim see *Mark Timmons,* "Necessitation and Justification in Kant's Ethics," Canadian Journal of Philosophy 22, 1992, pp. 377-398.

analytic, and since the conclusion follows from (1) and the other analytic premises, the conclusion is synthetic a priori. The master argument now proceeds as follows:

5. Necessarily, for any finite agent, she is susceptible to temptation to transgress duty out of self-love (from non-moral motives).
6. For any agent who is susceptible to temptation to transgress duty, that agent lacks a propensity to good. (From the definition of propensity to good.)
7. Either a person is good or evil (i. e., either a person's will is characterized by a propensity to good or by a propensity to evil). (Kant's rigorism.)

Therefore (from 5-7):

8. Necessarily, all finite agents (all human beings) have a propensity to evil.

This deduction, of course, is meant to show that human beings do not, by nature, have a good disposition, therefore, (given Kant's rigorism) they must have an evil disposition by nature. But given Allison's definition of a propensity to good, what this deduction shows (and all it shows) is that human beings are not holy beings. "For finite *holy* beings (who could never be tempted to violate duty) there would be no doctrine of virtue. . . ."[46] But, showing this seems to fall short of showing that all human beings have an evil disposition, in the sense of lacking a proper orientation of the will (at least if my analysis of this notion is correct). That is, there is a gap between the claim that a human being has an ineliminable susceptibility to act from motives of self-love, and the claim that, as a matter of fact, that human beings, in contexts involving moral requirements, at least sometimes fail to act solely from the motive of duty (where, of course, the failure is imputable). The essence (the necessary and sufficient conditions) of an evil disposition or radical evil is the failure of moral motives to have motivational dominance. It is motivational dominance, or lack of it, that is involved in the so-called degrees of radical evil: frailty, impurity, and wickedness. Not only is one *tempted* to act from motives of self-love, but as a matter of fact, such motives have (at least some of the time) the sort of dominance that should be possessed only by moral motives.

This point can be sharpened if we consider the notion of *susceptibility to temptation*. This notion (as it figures in Allison's deduction) involves essentially two components: (1) first, one freely takes non-moral considerations of self-love to be good reasons for action and thus is disposed to act for such reasons; and thus (2) one is at least *capable* of failing to give proper motivational dominance to moral requirements. But being capable of failing to give proper motivational dominance to moral requirements is one thing, and actually failing to do so (i. e., adopting as one's supreme maxim "occasional deviation" from the moral law is another. In fact, Allison calls attention to a parallel distinction in connection with Kant's conception of virtue: "Also central to Kant's conception of virtue

[46] Die Metaphysik der Sitten VI: 383, 188.

is the distinction between actual strength of character or self-control and the mere capacity *(Vermogen)* for it. The latter is possessed by all rational agents, no matter how weak or evil in virtue of their moral autonomy; the former must be acquired through a process of self-discipline."[47] Just as capacity for virtue is a necessary, but not a sufficient condition for virtue, so susceptibility to temptation is a necessary condition (one that holy wills lack) but not a sufficient condition for evil.

Allison himself worries that since his deduction really only amounts to showing the human beings are not holy wills, Kant's UT thesis, which seems rather remarkable, is thereby trivialized. He writes:

> Clearly what is needed at this point in order to put some bite back into the doctrine that there is a universal propensity to evil is a reason, apart from the general principle of rigorism, for regarding the lack of a propensity to good as equivalent to, or at least as entailing, an actual propensity to evil. Fortunately, although Kant never spells out his position with sufficient clarity, the basis for an explanation is provided by the previous analysis. The essential point is that the very fact that we only obey the law reluctantly *(ungern)* indicates not merely a lack of holiness but also an actual propensity to subordinate moral considerations to our needs as sensuous beings, that is, a tendency to let ourselves be tempted or "induced" by inclination to violate the moral law even while recognizing its authority.[48]

This response (to the sort of worry I have been raising) trades on an ambiguity in talk about an *actual propensity to subordinate moral considerations to our needs*. Taken in one sense, having such a propensity seems equivalent to merely being susceptible to temptation, in which case, again, the UT thesis simply comes to the claim that humans are not holy. Taken in another sense, talk of an actual propensity is equivalent to having chosen the sort of supreme maxim that characterizes an evil will: a maxim in which the incentives of morality and self-love are not properly ordered. The problem is that the claim: (1) finite rational agents are necessarily susceptible to temptation does not entail the claim: (2) finite rational agents necessarily have adopted an evil supreme maxim (and hence are guilty of being radically evil). Moreover I see nothing in Allison or Kant to bridge this gap. In conclusion, Allison's deduction of the Kant's UT fails to show that all human beings are guilty of radical evil in any interesting sense of the term.

This conclusion is reinforced by the fact that if Allison's argument is successful, then it is hard to see how we are responsible for our evil natures. It is intelligible

[47] *Allison* (op. cit, fn. 2), p. 164.

[48] Ibid., p. 157. It seems that Allison sometimes construes the notion of radical evil merely in terms of susceptibility to temptation (pp. 155ff, 159, 165, and 167), though at other times he seems to be construing radical evil much in the way I have as essentially involving the adoption by the agent of a certain sort of supreme maxim (pp. 147, 151, and 170). It is the first, weaker reading that is involved in Allison's deduction of the UT; the argument won't fly if one instead inserts the stronger reading of radical evil.

to claim that we are responsible for whatever supreme maxim we have (though the claim is not entirely trouble free). However, it does not seem intelligible to claim that we are responsible for our susceptibility to temptation. Susceptibility to temptation and certain 'natural' characteristics like eye color seem to be on a par.[49] But surely, any acceptable interpretation of Kant's doctrine must preserve Kant's claim that our evil nature can be imputed to us.

Of course, underlying Allison's deduction is a reading of good and evil disposition that differs from the one I have proposed. Since Allison's deduction exploits Kant's character rigorism, talk of a propensity to good is being construed as equivalent to having a good disposition. Contrary to how the notions of good and evil disposition were characterized above, this means that having a good disposition is equivalent to being a holy will, and all other wills, even the person of virtuous character, possesses an evil disposition. I have already made my case for analyzing these notions differently. But the issue here really isn't about how exactly to construe Kant's notions of good and evil disposition (and related notions). However we construe them, Kant faces a dilemma. If one construes the notion of radical evil as Allison does, then we have a deduction of UT, but this the thesis so understood fails to show anything interesting and, in particular, severs the connection between radical evil and imputation. On the other hand, if one construes the notion of radical evil as I have, then Allison's deduction fails, since there is a gap between claiming that one is necessarily susceptible to temptation and the claim that one has adopted an evil supreme maxim.[50]

VII. Conclusion

With some basic themes and distinctions pertaining to Kant's theory of action and practical reasoning on the table, we proceeded to explore in what sense morally evil actions are imputable, irrational, and evil. This provided a basis for an analysis of Kant's notion of radical evil and related notions (good and evil disposition), which we then employed in connection with questions about the relation between radical evil and moral worth. I argued that radical evil is

[49] This analogy was suggested to me by Nelson Potter.

[50] There is obviously much more to be said about Kant's doctrine of radical evil in general and the UT in particular. For instance, my discussion of Allison raises the question of just how the UT is to be understood. One interpretation is this: *All human beings, qua human, are radically evil* — an interpretation which apparently rules out the possibility of human beings ever being able to overcome (in this life) their evil natures. Another interpretation is this: *All human beings are born into a state of radical evil* — an interpretation that allows for the possibility that human beings can, in this life, overcome their evil natures. The former interpretation fits with Allison's story about radical evil, the latter fits with my own. For a more detailed investigation of these matters, see *Mark Timmons*, "Good Wills, Holy Wills, and Radical Evil," (manuscript in preparation).

compatible with the possibility of performing actions having moral worth. Finally, we turned to Kant's infamous claim that all human beings are necessarily evil and considered Allison's attempt to provide the missing deduction of this thesis. I argued that Allison's deduction does not work and that this thesis remains problematic.[51]

Zusammenfassung

Für Kant manifestiert sich das moralisch Böse in allen seinen Varianten — das Böse nämlich, das im Charakter einer Person wurzelt — in Handlungen, die auf der einen Seite vom eigenen Standpunkt des Handelnden her expliziert werden können und die deshalb zurechenbar sind, die aber auf der anderen Seite in einer gewissen Hinsicht als irrational bezeichnet werden müssen. Weil das Böse im Charakter der Person wurzelt, „verdirbt es den Grund aller Maximen" und verdient deshalb „das *radikal* Böse" genannt zu werden. Darüber hinaus sind menschliche Wesen nach Kant nicht nur für das Böse anfällig, es ist vielmehr unentrinnbar menschliches Schicksal, böse zu sein. Diese Thesen werfen eine Reihe von Fragen auf, unter ihnen die folgenden: (1) Wie läßt sich, Kants Auffassungen über die Natur menschlichen Handelns vorausgesetzt, die Möglichkeit irrationalen, aber gleichwohl vom Standpunkt des Handelnden her explizierbaren und also freien Handelns erklären? (2) Was ist das Wesen des radikal Bösen? (3) In welchem Sinne verdirbt es den Grund aller Maximen? (4) Warum behauptet Kant, das radikal Böse sei unentrinnbares Schicksal des Menschen? Der Beitrag ist diesen Fragen gewidmet. Er befaßt sich zuerst mit einigen Grundfragen und Grundunterscheidungen von Kants Handlungstheorie und seiner Theorie der praktischen Vernunft (Abschnitte I und II). Das erlaubt uns, Kants Auffassungen über das moralisch Böse zu verstehen. Abschnitt III wendet sich Kants Analyse des moralisch Bösen in seinen verschiedenen Ausprägungen zu, um auf dieser Grundlage eine allgemeine Charakterisierung des moralisch Bösen geben zu können (Abschnitt IV). Abschnitt V befaßt sich mit Kants Behauptung, das radikal Böse verderbe den Grund aller Maximen, und mit ihren Implikationen für die Frage nach dem moralischen Wert von Handlungen überhaupt. In Abschnitt VI geht es um das, was man „Kants Universalitätsthese" (UT) nennen kann, d.i. seine Behauptung, menschliche Wesen seien notwendigerweise radikal böse. In seinem jüngst erschienenen Werk über *Kant's Theory of Freedom* vertritt Henry Allison die Meinung, UT enthalte einen synthetischen Satz a priori, und er legt eine „Ableitung" von UT vor, was Kant nicht getan hat. Ich versuche zu zeigen, daß Allisons Ableitung nicht schlüssig ist, obwohl ich — leider — keine eigene Deduktion anbieten kann, die Kants These stützt, noch einen Weg sehe, Kants eigene Verteidigung von UT mit ihrem angeblich apriorischen Status in Übereinstimmung zu bringen. Kants UT bleibt problematisch.

[51] I wish to thank Nelson Potter for his comments on an earlier draft of this paper.

MARY TRIGGS
DEPARTMENT OF PHILOSOPHY

Jahrbuch für Recht und Ethik

Annual Review of Law and Ethics

Herausgegeben von

B. Sharon Byrd · Joachim Hruschka · Jan C. Joerden

Band 2

Duncker & Humblot · Berlin

Jahrbuch für Recht und Ethik

Annual Review of Law and Ethics

Band 2 (1994)

Themenschwerpunkt:

Zurechnung von Verhalten
Imputation of Conduct

Herausgegeben von

B. Sharon Byrd
Joachim Hruschka
Jan C. Joerden

Duncker & Humblot · Berlin