

## MORALITY AND FREEDOM: KANT'S RECIPROCITY THESIS

Henry E. Allison

At the end of the second part of the *Groundwork* Kant reflects that he has so far shown only that the autonomy of the will ("the property the will has of being a law to itself")<sup>1</sup> is the supreme principle of morality in the sense of being the ultimate presupposition of morality as it is commonly understood.<sup>2</sup> The articulation of this principle marks the culmination of an analytic or regressive argument, the aim of which is to uncover and present with philosophical precision the basic assumptions and principles of the ordinary, pre-philosophical conception of morality. Although Kant hardly minimizes this result, he also displays an awareness of the fact that, of itself, it leaves unanswered the crucial question of validation. Accordingly, he sets himself the task in the third part of the *Groundwork* of answering the moral skeptic by showing that morality is not a "chimerical Idea," a mere "phantom of the brain."<sup>3</sup> In Kantian terms, this requires that he provide a transcendental deduction of morality.

Unfortunately, not only has Kant's effort to accomplish this important goal been severely criticized by even his most sympathetic critics, but the purely exegetical question of what kind of argument, if any, the text supplies, has been the topic of an ongoing dispute. In fact, there seems to be no agreement as to whether the deduction is of the moral law, the categorical imperative, freedom, all three; or even whether there is properly a

---

<sup>1</sup>*Groundwork of the Metaphysics of Morals*, *Kants gesammelte Schriften*, Berlin: Königliche Preussische Akademie der Wissenschaften, 1901–1922, IV, 440.

<sup>2</sup>For the distinction between two senses in which Kant speaks of a "supreme principle of morality," one of which applies to the categorical imperative and the other to the principle of autonomy, see Lewis White Beck, *A Commentary on Kant's Critique of Practical Reason* (Chicago: University of Chicago Press, 1960), p. 122 and T. C. Williams, *The Concept of the Categorical Imperative* (Oxford: Clarendon Press, 1968), pp. 33–35.

<sup>3</sup>*Kants gesammelte Schriften*, IV, 445.

deduction at all.<sup>4</sup> Furthermore, the uncertainty about the *Groundwork* has given rise to additional questions about its relationship to the *Critique of Practical Reason*. The problem here is that while in the *Groundwork* Kant at least seems to have attempted a transcendental deduction of the moral law and/or the categorical imperative on the basis of the necessity of presupposing the Idea of freedom, in the *Critique of Practical Reason* he explicitly denies the possibility of any such deduction and claims instead that the moral law as a “fact of reason” can serve as the basis for a deduction of freedom.<sup>5</sup> This suggests a significant reversal in Kant’s thought regarding the justification of morality; although even here the existence of such a reversal has been denied both by those who see no real deduction of the moral law in the *Groundwork* and by those who claim to find a deduction in the *Critique of Practical Reason* as well as in the *Groundwork*.<sup>6</sup>

Undoubtedly, most of the confusion can be traced to Kant’s own confusing and sloppy formulations of both his problematic and his argument, especially in *Groundwork III*. Nevertheless, I do believe that part of the blame can be attributed to a failure on the part of Kant’s critics to give proper attention to a thesis which is at least relatively clear, and which looms large in both the *Groundwork* and the *Critique of Practical Reason*. This is the claim that freedom of the will and the moral law are reciprocal concepts. Kant affirms this explicitly in both works; correlatively, he also insists in both works that, although the moral law (or, better, the bindingness of the moral law for all rational agents) expresses a synthetic *a priori* proposition, it would be analytic if freedom of the will were presupposed.<sup>7</sup> For convenience sake I shall henceforth refer to this as the “Reciprocity Thesis.” Its significance stems from the fact that it

---

<sup>4</sup>The fullest and most significant discussion of this issue is by Dieter Henrich, “Die Deduktion des Sittengesetzes,” in *Denken im Schatten des Nihilismus*, ed. Alexander Schwan (Darmstadt: Wissenschaftliche Buchgesellschaft, 1975), pp. 55–112.

<sup>5</sup>*Kants gesammelte Schriften*, V, 42–50.

<sup>6</sup>For a recent discussion of the “Reversal Thesis,” which includes an excellent account of the whole debate, see Karl Ameriks, “Kant’s Deduction of Freedom and Morality,” *Journal of the History of Philosophy* 19 (1981), pp. 53–79.

<sup>7</sup>See *Kants gesammelte Schriften*, IV, 447; V 29–31.

entails that freedom of the will is not only a necessary but also a sufficient condition of the moral law.<sup>8</sup> Clearly, this thesis underlies Kant's attempt in the *Groundwork* to argue from freedom (or at least from the necessity of the presupposition of freedom) to the moral law, and in the *Critique of Practical Reason* from the moral law (as a putative "fact of reason") to the reality of freedom.

The goal of this paper is to provide a defense of this thesis, which lies at the very heart of Kant's moral philosophy. The defense will begin with an examination of the Kantian texts; but since his "official" arguments for the thesis are obviously inadequate, it will be necessary to go considerably beyond Kant's explicit statements on the topic. Thus, the proposed defense is also a reconstruction of Kant's argument, albeit one based largely on material which Kant himself has provided. Since the Reciprocity Thesis is only the first step in the Kantian justification of morality, a defense of this thesis will not amount to a complete defense of the Kantian "deduction." It is, however, a necessary first stage in such a project. Moreover, I hope to show that the recognition of the cogency and systematic role of the thesis is itself enough to obviate some of the standard objections to Kant's procedure. I also hope to show that, properly construed, the Reciprocity Thesis is not open to the devastating criticism which is frequently raised against it: namely, that it entails that no free action can be morally wrong.

## I

The best known and most perplexing of Kant's formulations of the Reciprocity Thesis is at the beginning of *Groundwork III*. After defining will [*Wille*] as a "kind of causality belonging to living beings so far as they are rational," and freedom (negatively construed) as "the property this causality has of being able to work independently of determination by alien causes," Kant offers his positive conception of freedom, which presumably "springs" from this negative one:

The concept of causality carries with it that of *laws* (*Gesetze*) in accordance with which, because of something we call a cause, something

---

<sup>8</sup>This point is noted by Henrich, "Die Deduktion des Sittengesetzes," pp. 89–90.

else—namely, its effect—must be posited (*gesetzt*). Hence freedom of will, although it is not the property of conforming to laws of nature, is not for this reason lawless: it must rather be a causality conforming to immutable laws though of a special kind; for otherwise a free will would be self-contradictory. Natural necessity, as we have seen, is a heteronomy of efficient causes; for every effect is possible only in conformity with the law that something else determines the efficient cause to causal action. What else then can freedom of will be but autonomy—that is, the property which will has of being a law to itself? The proposition “Will is in all its actions a law to itself” expresses, however, only the principle of acting on no maxim other than one which can have for its object itself and at the same time a universal law. This is precisely the formula of the categorical imperative and the principle of morality. Thus a free will and a will under moral laws are one and the same.<sup>9</sup>

Kant also argues for the same thesis in §6 of the *Critique of Practical Reason*. After contending on the basis of an analysis of the concept of a practical law 1), that such a law must be formal in the sense that it could only impose the formal condition of lawfulness on the maxims of a rational agent §4 and that only a will that is free in the transcendental sense could have its “determining ground” in such a law §5; he then §6 poses the problem: “Granted that a will is free, find the law which alone is competent to determine it necessarily” [*welches ihn allein nothwendig zur bestimmen tauglich ist*]. The proposed solution exploits the dichotomy between the form (lawfulness or universality) and the matter (desired object or end) of a practical principle developed in §§2–4. Kant claims that since 1) a free will (by definition) must be independent of all “empirical conditions,” which includes the “material” element of practical principles; and that 2) a free will must nonetheless be “determinable” (presumably according to some law); that 3) “the legislative form, insofar as it is contained in the maxim, is the only thing which can constitute a determining ground of the [free] will.” On this basis Kant concludes at the very beginning of the Remark following the analysis that “freedom and unconditional practical law reciprocally imply each other.”<sup>10</sup>

The argument at this point is completely hypothetical, and con-

<sup>9</sup>*Kants gesammelte Schriften*, IV, 446. *Groundwork of the Metaphysics of Morals*, Eng. trans. H. J. Paton (New York: Harper and Row, 1956), p. 114.

<sup>10</sup>*Kants gesammelte Schriften*, V, 29.

sequently does not involve any claims concerning the reality of either freedom or an unconditional practical law. Nevertheless, given the identification §7 of an unconditional practical law with the moral law ("So act that the maxim of your will could always hold at the same time as a principle establishing universal law"),<sup>11</sup> it is but a short step to the conclusion that "It [the moral law] would be analytic if freedom of the will were presupposed." Admittedly, this last claim is somewhat strange. How, one might ask, could the presupposition of freedom convert a synthetic proposition into one that is analytic? The most reasonable reading, I take it, is that analyticity is to be attributed to the hypothetical, "If freedom then the moral law," and to its reciprocal. Kant clarifies his position near the end of the *Analytic of Pure Practical Reason* when he remarks:

if [*per impossibile*] we saw the possibility of freedom of an efficient cause, we would see not only the possibility but also the necessity of the moral law as the supreme practical law of rational beings, to whom freedom of the causality of their will is ascribed. This is because the two concepts are so inextricably bound together that practical freedom could be defined through the will's independence of everything except the moral law.<sup>12</sup>

Kant does not explicitly argue for the Reciprocity Thesis in *Religion within the Limits of Reason Alone*. Nevertheless, in the course of denying that the source of moral evil can be located either in man's sensuous nature or in his "morally legislative reason" [*Wille*], he does suggest that to affirm the latter is equivalent to saying that "reason could destroy the authority of the very law which is its own, or deny the obligation arising therefrom." This, however, he claims is impossible because:

To conceive of oneself as a freely acting being and yet as exempt from the law which is appropriate [*angemessen*] to such a being (the moral law) would be tantamount to conceiving a cause operating without any laws whatsoever (for determination according to natural laws is excluded by the fact of freedom); this is a self-contradiction.<sup>13</sup>

---

<sup>11</sup>Ibid., p. 30.

<sup>12</sup>Ibid., p. 93–94.

<sup>13</sup>*Kants gesammelte Schriften*, VI, 35. *Religion within the Limits of Reason Alone*, Eng. trans. T. M. Greene ed., Hoyt Hudson (New York: Harper & Row, 1960), p. 30. See also *Reflexion* 7202, XIX: 281.

The argument in the *Groundwork*, in particular, appears to be vitiated by a gross equivocation regarding the concept of law. As even Paton, the most sympathetic of Kant interpreters notes, it is hardly legitimate to jump (as Kant there seems to do) from the notion of a causal law, which is a law connecting causes and effects, to a “law of freedom,” which, by definition, would be a law for decision itself, not one which connects decisions (as causes) with their effects in the phenomenal world.<sup>14</sup> Leaving this aside, however, it is possible to specify a common core of argumentation that is contained, implicitly at least, in all of the texts. The argument takes roughly the following form: 1) As a “kind of causality” the will must, in some sense, be law-governed or, in the language of the Second Critique, “determinable” according to some law (a lawless will is an absurdity). 2) As free, it cannot be governed by laws of nature. 3) It must, therefore, be governed by laws of a different sort; that is, self-imposed ones. 4) The moral law is the required self-imposed law.

Although a compatibilist would certainly object to steps 2 and 3, the major difficulties which we need consider concern steps 1 and 4. Clearly, if a free will (in a non-compatibilist sense) is to be law-governed or “determinable,” it can only be through self-imposed laws. In that minimal sense, then, the positive concept of freedom (autonomy) can be derived analytically from the negative concept (independence). However, apart from the already noted equivocation regarding the concept of law, Kant does not seem to offer any argument in support of the claim that a free will must be law-governed or “determinable” at all. On the contrary, the account in the Second Critique suggests that this essential question is simply begged.

At first glance at least, step 4 appears to be equally problematic; for even if we assume that a free will must be governed or determinable by a self-imposed law, it does not seem at all obvious that only the moral law, as defined by Kant, can do the job. In fact, considering only the *Groundwork* account, it once again seems that Kant has begged the main question by means of his prior characterization of the principle of autonomy as the “supreme principle

---

<sup>14</sup>H. J. Paton, *The Categorical Imperative* (London: Hutchinson & Co., 1958), p. 211. A similar view is expressed by Sir David Ross, *Kant's Ethical Theory* (Oxford: Clarendon Press, 1954), pp. 70–71.

of morality.” This characterization makes it all too easy for Kant to slide from the claim that a free will is autonomous in the sense that it is determinable only by self-imposed laws to the claim that the law which it spontaneously yet necessarily imposes upon itself is the moral law.<sup>15</sup>

With regard to the latter problem, it is crucial to note that Kant holds that the moral law is the only conceivable candidate for a practical law. Consequently, for Kant at least, the claims that a free will is necessarily subject to a practical law (step 1) and that it is necessarily subject to the moral law (step 4) are equivalent. According to Kant's implicit definition, a practical law is an objectively and unconditionally valid practical principle. To claim that a practical principle is *objectively* valid is to claim that it holds for all rational agents, whether or not they in fact adhere to it, that is, whether or not it holds subjectively (as a maxim). In the case of imperfectly rational beings such as ourselves, such a principle takes the form of an imperative (which is likewise objectively valid). The imperative is hypothetical if its objectivity is a function of certain ends or desires; it is categorical if this is not the case. An objectively valid practical principle is also *unconditionally* valid just in case it holds independently of any ends or desires. The imperative issuing from such a principle is always categorical, whatever its grammatical form.<sup>16</sup>

---

<sup>15</sup>Rüdiger Bittner, *Moralisches Gebot oder Autonomie* (Freiburg/Munich: Verlag Karl Alper, 1983), pp. 119–134, claims that the argument of the third part of the *Groundwork*, particularly the Reciprocity Thesis, is vitiated by this slide, which is, in turn, based on a confusion of two senses of ‘autonomy’. A similar line of objection is also developed by Gerold Prauss, *Kant über Freiheit als Autonomie* (Frankfurt: Vittorio Klostermann, 1983). Prauss insists upon the need for distinguishing between a morally neutral and a specifically moral sense of ‘autonomy’ and criticizes Kant for a failure to be clear on this point.

<sup>16</sup>I have claimed that the above account is based on Kant's implicit definitions because his official definitions of these notions are notoriously confusing. Thus, in some places he simply identifies a practical law with an objectively valid principle (for example, *Critique of Practical Reason* §1, *Kants gesammelte Schriften*, V, 19); while in others he seems to regard practical laws as constituting a subset of objectively valid practical principles. (for example, *Kants gesammelte Schriften* V, 20 and *Groundwork*, *Kants gesammelte Schriften*, IV, 416, 420). This, in turn, is connected with Kant's equally notorious confusion of laws and imperatives. Since all imperatives are objectively valid (in contrast to maxims) I take the latter position to be the one to which Kant is committed. For a further account of some of these issues, see Beck, *A Commentary*, pp. 79–84 and 121–122.

Given this conception, Kant claims 1) that such a law must be “formal,” since it abstracts from all ends or desires (which constitute the “matter” of a principle), and 2) that, as such, it can require only that rational agents select their maxims on the basis of their suitability as universal laws. This is, of course, precisely what the moral law or, better, the categorical imperative requires. Although much more work would be needed to make this line of argument fully convincing, I do find it a plausible unpacking of the implications of Kant’s definitions. In any event, for the purposes of this paper, I propose to accept this claim. Thus, its goal will have been achieved if, in Kant’s words, it can be shown that “freedom and unconditional practical law reciprocally imply each other.”

## II

Kant’s claim that the notion of a lawless will involves an absurdity places him squarely within the metaphysical tradition that rejects the conception of a “liberty of indifference.” This rejection is a constant in Kant’s thought; it can be found in his earliest significant discussion of freedom, where he defends the Leibnizian view.<sup>17</sup> It resurfaces, however, in the “critical period” in connection with a very different, radically un-Leibnizian, conception of the will and its freedom.

The gist of this new conception of the will is indicated in the famous statement in the *Groundwork* that “Everything in nature works in accordance with laws. Only a rational being has the power to act in accordance with his *Idea* of laws—in accordance with principles—and only so has he a will.”<sup>18</sup> Kant then goes on to define the will as practical reason on the grounds that reason is required to derive actions from laws.<sup>19</sup> Somewhat later he defines “will” [*Wille*] as a “kind of causality belonging to rational beings so far as they are rational.”<sup>20</sup> Rationality, construed as the capacity to form general principles, together with the power to act on the basis of these principles, thereby producing changes in the phenomenal

---

<sup>17</sup>*A New Exposition of the First Principles of Metaphysical Knowledge*, Proposition IX. *Kants gesammelte Schriften*, I, 398–405.

<sup>18</sup>*Kants gesammelte Schriften*, IV, 412.

<sup>19</sup>*Ibid.*

<sup>20</sup>*Ibid.*, p. 446.



world (if only in the psychological state of the agent) are, therefore, the defining characteristics of "will" as Kant construes it in the *Groundwork*. Only a being with both of these capacities can be said to have a will. Such a being is also one for whom reason is practical.

If one is to understand Kant's thought at this point, it is crucial to realize that "rationality" is here construed in a very broad sense. Since all that is required is the capacity to form and act upon general principles, an agent is "rational" in the relevant sense even when the principles he adopts as rules for action are morally pernicious, imprudent, or even self-defeating, that is, "irrational" in the usual sense. Kant's technical term for the "principle" or "Idea of law" on the basis of which rational agents supposedly act is 'maxim'. Consequently, it is appropriate to begin our examination of Kant's claim that a lawless will is an absurdity with a brief consideration of his account of maxims. This consideration must, of necessity, be superficial. It will concentrate solely on the presumed role of maxims in human action, thereby ignoring many of the complexities and ambiguities of Kant's account, as well as the interesting questions regarding the specification of maxims.

As is all too frequently the case, in his characterization of maxims Kant succeeds in being technical without being precise. For present purposes, however, it suffices to describe a maxim as a subjective practical principle, that is, a general rule or policy on which a rational agent actually acts in a given situation and tends to act in relevantly similar situations.<sup>21</sup> Expressed algebraically, maxims have the form: "To do A if B." As subjective, maxims are closely connected with the "interests" of an agent, which are themselves never the simple result of mere impulse or sensuous desire, but always involve some conception of an end. It is only because I consciously choose to pursue certain ends or, equivalently, have certain interests, that I adopt certain policies of action, designed to realize these ends. A maxim thus has a purposive component built into it; although this component need not be made explicit in the formulation.<sup>22</sup> Moreover, this is true even when the interest is

---

<sup>21</sup>*Kants gesammelte Schriften*, IV, 400n and 420n; V, 19; and VI, 225.

<sup>22</sup>For a useful discussion of this point and, indeed, of the whole topic of Kant's view of maxims see Onora Nell, *Acting on Principle* (New York and London: Columbia University Press, 1975), pp. 34–42.

“pure,” that is, not based on any sensuous desire for the object, as is supposedly the case in action for the sake of the moral law.

Although there can be no quarrel with the claim that people often act on the basis of consciously adopted maxims, it is also frequently thought that the emphasis Kant places on maxims in his account of human action makes human behavior appear much more rule-governed than it actually is.<sup>23</sup> This is particularly true if maxims are construed as relatively fixed policies of life or *Lebensregeln*, which specify the most fundamental choices of an individual, and which, as such, are contrasted with mere precepts or “rules of thumb.”<sup>24</sup> So construed, maxims certainly provide ready candidates for moral evaluation, and it is to maxims in this sense that Kant appeals in his well known attempts to illustrate the application of the categorical imperative. The problem is simply that many, if not most, human actions cannot be plausibly regarded as the result of an explicit reflection on rules of this sort; but this neither exempts them from moral evaluation (we are justly condemned for our “impulsive” acts), nor reduces them to mere bits of behavior, not worthy of being termed “actions.”

There are, I think, two possible responses to this fairly obvious line of objection, both of which have a basis in the Kantian texts. The first involves a certain broadening of the notion of a maxim, making it roughly equivalent to an intention.<sup>25</sup> On this interpretation, the claim that an agent acted on the basis of a maxim does not entail either that he acted on the basis of a principle to which he has been committed for any determinate length of time or that he explicitly “subsumed” his action under this principle, in the manner of someone who goes through all of the steps of an Aristotelian

---

<sup>23</sup>This, again, is a fairly common line of criticism. A good formulation of it is provided by Marcus Singer, *Generalization in Ethics* (New York: Alfred A. Knopf, 1961), pp. 245–46.

<sup>24</sup>On this point see Beck, *A Commentary*, p. 78, and Rüdiger Bittner, “Maximen,” *Akten des 4. Internationalen Kant-Kongresses*, Mainz, 1974, ed. G. Funke and J. Kopper (Berlin: de Gruyter, 1974), pp. 485–498.

<sup>25</sup>In *Eine Vorlesung über Ethik*, ed. Paul Menzer (Berlin: Rolf Hesse, 1924), pp. 52–53, Kant refers to the universalizability of the intention (*Intention*) of an action, thereby treating intentions much as he later does maxims. The problem is complicated for the English reader, however, by the fact that, in his translation of the *Critique of Practical Reason*, Beck frequently renders ‘*Gesinnung*’, as ‘intention’.

practical syllogism. It entails only that he acted with conscious intent, that there is a specifiable reason for the action. To formulate the maxim is to describe this intent and to give the reason. Here it will be helpful to follow Onora Nell, who, appealing to Kant's later formulation in the *Metaphysics of Morals* (which presupposes the *Wille-Willkür* distinction), characterizes a maxim as a "determination of the power of choice" [*Willkür*]. As she correctly points out in her comment on this characterization, "To say that an agent's power of choice is determined is simply to say that he intends to do a specific sort of act or pursue a specific end in some situation. If an agent has a maxim 'To do A if B', then he intends to do A if B."<sup>26</sup> All that needs to be added at this point is that the converse of the last claim likewise holds. If an agent really intends to do A if B, then he has a maxim "To do A if B," whether or not he is explicitly aware of it.

By construing maxims in this way it is possible to ascribe them to many actions which are performed "on the spur of the moment," without reflection or the explicit adoption of a settled policy. To borrow an example from Onora Nell: A person can suddenly decide to have an extra cup of coffee one morning without any deliberation and without the adoption of a specific policy regarding the amount of coffee to be consumed each morning. Certainly, we cannot claim plausibly that such an action involves a maxim in the sense of a *Lebensregel* (or even a "rule of thumb"). Nevertheless, we can connect it with a maxim in the broad sense insofar as we can attribute an intention to the agent, for example, to get warm or to combat the effects of a sleepless night. Moreover, as Nell notes, even if the agent himself does not reflectively formulate this maxim or intention, it can still (in principle) be discovered by determining what changes in the circumstances would have led him to decline the extra cup.<sup>27</sup>

This line of interpretation suggests that a Kantian maxim is very close, if not equivalent, to the "plan as it were" that J. L. Austin claims to be an essential ingredient in an intentional action. According to Austin, for an agent to act intentionally, which he is careful to distinguish from acting deliberately or purposefully, he

---

<sup>26</sup>Onora Nell, *Acting on Principle*, p. 40.

<sup>27</sup>*Ibid.*, p. 41.

must have a conscious idea of what he is doing, and this requires having, at least in some minimal sense, a plan of action.<sup>28</sup> To have such a plan, for example, “Do A if B,” is precisely what it means to “know what one is doing,” while the latter is a necessary condition of an intentional action. Following Austin, it can, therefore, be claimed that for any description under which an action is intentional, it must be possible to assign some “plan as it were” to the agent. But the same can be said, *mutatis mutandis*, of maxims broadly construed. Moreover, since the notion of a will without any intentions is manifestly absurd, it follows that the notion of a will without maxims (a “lawless will”) is likewise absurd.

Although there are good reasons for interpreting maxims in this way (How else can one square the fact that people are deemed morally responsible for their so-called “impulsive” acts with Kant’s insistence that morality is concerned with the maxims of action?), it is not necessary to insist upon it here. Even if we assume that by ‘maxims’ is meant something like *Lebensregeln*, which would preclude any straightforward identification of intentional action with action based on a maxim, it can still be maintained that an agent capable of intentional action at all, that is, one with the capacity for rational choice (in the broad sense of ‘rational’) cannot be totally without maxims. The point here is simply that an agent completely bereft of maxims (in the sense of *Lebensregeln*) would also be without any self-chosen goals or interests, and this means that he would have no basis for rational choice. Consequently, his “actions” would have to be regarded either as random happenings (which is absurd) or as direct responses to stimuli, explicable in neurophysiological terms. In short, his “actions” would, like other natural occurrences, be “in accordance with laws,” not, as in the case of rational agents, “in accordance with the Idea of laws.”

### III

The preceding analysis may help to explain and give plausibility to Kant’s claim that rational agents act in accordance with the “Idea of laws,” but it obviously does not suffice to establish the thesis that a free will (the will of a rational agent) must be law-governed in any

---

<sup>28</sup>J. L. Austin, “Three Ways of Spilling Ink,” *The Philosophical Review* 75 (1966), pp. 427–440, esp. 437–438.

but a trivial sense of 'law'. In fact, we have seen that Kant defines maxims as subjective practical principles and explicitly contrasts them with objective practical principles or laws. Thus, given Kant's own definitions, there can be no immediate transition from being maxim-governed to being subject to an "unconditional practical law." We are still in need of an argument to bridge this gap.

The argument that comes immediately to mind at this point is a familiar one, and so is the objection to it. I do not believe that it is Kant's own argument (although it is frequently taken to be such), but I do think that it can be construed as an essential first step in an extended Kantian argument. I also think that when it is viewed merely as a first step in an extended argument rather than as a complete argument in its own right, the standard objection loses its force.

In the endeavor to sketch this argument or, more accurately, argument-segment, it will be helpful to return to the analysis of maxims in terms of intentions. This analysis strongly suggests that to stipulate an agent's maxim in performing a certain action is to give the agent's "reason" for that action, at least in one important sense of that notoriously elusive notion. (To state my intention *in* X-ing—the "plan as it were" that I have "in mind" is to give my reason *for* X-ing.) More specifically, it is to give the kind of reason in terms of which an action can be justified (or criticized) as opposed to being explained or even excused.

Such justification (or criticism) certainly includes, but is not limited to, the moral variety. A given action could be praised as morally appropriate or as prudent or, correlatively, condemned as immoral or as foolish. These are obviously quite different kinds of evaluations, but the key point is that in all cases they are based upon assumptions regarding the agent's intention to act in a certain way in a given set of circumstances. Moreover, in both the moral and prudential contexts the justification takes a similar form: namely, showing that the reason (in the sense of intention) for acting in a certain way is a "good reason." Naturally, the same can be said, *mutatis mutandis*, regarding the criticism of actions, whether this be on moral or on prudential grounds.

The next step is to note that in claiming that one's reason for acting in a certain way is a "good" in the sense of justifying reason, one is, implicitly, at least, assuming its appropriateness for all rational beings. The intuition behind this is simply that if reason R

justifies my X-ing in circumstances C, then it must also justify the X-ing of any other agent in such circumstances. As Marcus Singer, paraphrasing Sidgwick, remarks, "A reason in one case is a reason in all cases—or else it is not a reason at all."<sup>29</sup> To be sure, there is a perfectly legitimate sense in which I might claim that something is "right for me" and not for others; but this must be construed as an elliptical way of stating that there is something peculiar about my circumstances (which can include, among other things, my desires and capacities). Thus, I might claim that a course of action, say going to the race track to relax, is justifiable for me because of my superior ability as a handicapper, great wealth, or luck, etc., while it is not justifiable for others who lack these attributes. What I may not do is to claim that the possession of these attributes justifies my action but not that of other similarly inclined and endowed agents. In roughly this way, then, the universalizability of one's intention, maxim or plan of action, seems to be presupposed as a condition of the possibility of justifying one's action, even when this justification does not take an explicitly moral form.

Finally, a rational agent cannot simply refuse to play the justification game, that is, refuse to concern himself with the question of whether the reasons for his actions are "good" reasons, at least in a non-moral sense of "good." This is, of course, not to say that such an agent always acts on the basis of good and sufficient reasons or that, in retrospect, he must always believe himself to have done so. The point is rather the familiar one that an agent for whom the whole question of justification is irrelevant, who never weighs the reason for his action, who acts without at least believing at the time that his reasons are "good" reasons, would not be regarded as rational. But since, as we have just seen, to regard one's reason for acting in a certain way as "good" is to assume its legitimacy for all rational beings in similar circumstances, it would seem, so the argument goes, that a rational agent cannot reject the universalizability test without, at the same time, denying his rationality. This, in turn, means that the universalizability test functions as the ultimate standard governing one's choice of maxims or, equivalently, that it has the status of a practical law.

This line of argument is too familiar to require further elabora-

---

<sup>29</sup>Marcus Singer, *Generalizations in Ethics*, p. 57.

tion, and so, too, is the objection to it. The problem is simply that one cannot move from the claim that every rational agent must regard his principles of action as universalizable in the sense that he be willing to acknowledge that it would be reasonable for every other agent in the relevantly similar circumstances<sup>30</sup> to adopt the same principles, or even that such agents ought to adopt them (where the "ought" is the ought of rationality), to the desired conclusion that the agent ought to be able to will (as a universal law) that every rational agent act on the basis of the principle in question. The rational egoist might very well be willing to admit that the maxims on which he acts in pursuit of his own perceived self-interest are also those on which every other rational agent ought to act (and would act, if sufficiently enlightened). It hardly follows from this, however, that the rational egoist is committed (on pain of self-contradiction) to will that all other rational agents behave likewise.<sup>31</sup>

While there can be little question about the cogency of this line of objection, considered as a response to the project of somehow deducing morality, conceived in Kantian terms, from the concept of rationality, there are serious questions about its relevance to Kant's own procedure. The reason that this is generally thought to provide a decisive criticism of Kant can no doubt be attributed to Kant's misleading claim that, since moral laws hold for rational beings as such, they ought to be derived from the "general concept of a rational being as such."<sup>32</sup> This is intended, however, to preclude any appeal to anthropology, that is, to empirical knowledge of human nature, and not to suggest that the reality of moral

---

<sup>30</sup>I am obviously ignoring the whole problem of specifying "relevantly similar circumstances," which is a critical issue in its own right. I think, however, that I am here in agreement with the analysis provided by Singer, *Generalizations in Ethics*, pp. 17–33.

<sup>31</sup>Concise versions of this criticism are given by A. W. Wood, "Kant on the Rationality of Morals," *Proceedings of the Ottawa Congress on Kant in the Anglo-American and Continental Traditions* Held October 10–14, 1974, edited by P. Laberge, F. Duchesneau, B. C. Morrissey (Ottawa: The University of Ottawa Press), 1976, pp. 94–109; and by Gilbert Harman, *The Nature of Morality* (New York: Oxford University Press, 1977), pp. 76–77. In Harman's case, however, there is absolutely no attempt to connect the criticism to the Kantian texts.

<sup>32</sup>*Groundwork of the Metaphysics of Morals*, *Kants gesammelte Schriften*, IV, 412.

obligation can be deduced from the “mere concept” of a rational being. In fact, not only does Kant himself not attempt to deduce the moral law from this concept, he explicitly rejects the possibility of doing so. We must keep in mind that the starting point of Kant’s analysis is not the concept of a rational being *simpliciter*, it is rather the concept of a rational being possessing a free will (in the transcendental sense). This is because Kant realized that, for all that we can learn from its “mere concept,” practical reason might involve nothing more than the capacity to determine the best possible means for the satisfaction of one’s desires. Certainly many distinguished philosophers have thought as much; and there is nothing self-contradictory or otherwise absurd in the claim. Indeed, as Kant himself remarks in a highly significant but strangely neglected note in *Religion within the Limits of Reason Alone*:

For from the fact that a being has reason it by no means follows that this reason, by the mere representing of the fitness of its maxims to be laid down as universal laws, is thereby rendered capable of determining the will unconditionally, so as to be “practical” of itself; at least, not so far as we can see. The most rational mortal being in the world might still stand in need of certain incentives, originating in objects of desire, to determine his choice. He might, indeed, bestow the most rational reflection on all that concerns not only the greatest sum of these incentives in him but also the means of attaining the end thereby determined, without ever suspecting the possibility of such a thing as the absolutely imperative moral law which proclaims that it is itself an incentive and, indeed, the highest.<sup>33</sup>

The conclusion to be drawn from this is that the problem with the argument “from rationality” sketched above is not that it is totally wrongheaded, but merely that it is incomplete.<sup>34</sup> Let us see, then, if

---

<sup>33</sup>*Kants gesammelte Schriften*, VI, 126 note. Eng. trans. p. 21.

<sup>34</sup>One of the few recent commentators to grasp this point is Thomas E. Hill, Jr., “The Hypothetical Imperative,” *The Philosophical Review* 82 (1973), pp. 429–450. In discussing the syntheticity of the principle which Kant attempts to justify in *Groundwork* III, Hill correctly notes that subjection to the moral law cannot be derived analytically from the concept of a rational person, but that the freedom of the person plays an essential role in the argument. In fact, Hill states clearly that Kant’s argument rests on two poles. 1) The claim that any person that is negatively free is also positively free (which is supposed to be a matter of conceptual analysis). 2) The claim that every rational being is also negatively free (which is not a



we can meet with more success in the endeavor to establish the Reciprocity Thesis if we focus explicitly on the transcendental freedom as well as the rationality of the agent. As a first step in this process we shall take a brief look at the relevant aspects of Kant's account of freedom.

## IV

Central to Kant's conception of freedom is the contrast between practical and transcendental freedom. For present purposes, practical freedom (*freie Willkür, arbitrium liberum*) can be equated with the previously considered capacity of a rational agent to act on the basis of maxims, that is, in light of the "idea" or "representation" of a law. This involves the capacity to act independently of, and even contrary to, any particular desire. Instead of responding automatically to the strongest desire (the mark of a pathologically necessitated *Willkür* or *arbitrium brutum*), a practically free agent can weigh and evaluate his desires, give priority to some and suppress others. Only *qua* conceptually determined, for example, taken up or "incorporated into a maxim," does a desire constitute a reason for acting. This does not rule out the possibility of what we normally regard as impulsive behavior, for example, an action out of anger. The point is only that in such cases the agent must be thought to give into the emotion, to let it move him to action.<sup>35</sup>

---

matter of conceptual analysis). I am in complete agreement with Hill regarding the structure of Kant's argument in *Groundwork* III. My concern here is to provide an argument in support of the first of these two poles, that is, to provide the required conceptual analysis, which is something that Hill does not attempt to do.

<sup>35</sup>Admittedly, the above account of Kant's First Critique theory of practical freedom is grossly oversimplified. In particular, it ignores the fact that Kant explicitly connects such freedom with the capacity to act on the basis of imperatives (for example, A534/B562, A547–48/B575–76, A802/B830). This, in turn, has led many commentators to assume that, even here, Kant understands freedom in explicitly moral terms. In reality, however, in his First Critique account Kant does not focus exclusively, or even primarily, on moral imperatives. He is rather concerned with the presentation of the outlines of a general theory of rational agency. For a fuller account of my views on this topic see "Practical and Transcendental Freedom in the *Critique of Pure Reason*," *Kant-Studien* 93 (1982), pp. 271–290, and *Kant's Transcendental Idealism: An Interpretation and Defense* (New Haven and London: Yale University Press, 1983), Chapter 15.

Transcendental freedom, by contrast, is usually defined as absolute spontaneity or as complete independence from any determination by antecedent conditions.<sup>36</sup> This creates the impression that the difference between practical and transcendental freedom is between a modest conception, presumably one that a compatibilist might accept, and a radical conception, requiring indeterminism together with all of its well known difficulties. After all, independence from “pathological necessitation” is hardly equivalent to independence from all causal determination; and if practical freedom involves only the former, then it is far from obvious that it requires indeterminism. Nevertheless, Kant seems to maintain that it does. At least he claims in the *Critique of Pure Reason* that “the practical concept of freedom is based on this *transcendental Idea*,” (A533/B561) and even that “The denial of transcendental freedom must, therefore, involve the elimination of all practical freedom” (A534/B56).

Largely as a result of passages such as these, Kant is frequently deemed guilty of an illicit slide from a respectable conception of practical freedom (pathological independence) to a disreputable or, at best, highly problematic transcendental conception. It is sometimes further claimed that the latter conception brings with it no discernible advantages and many significant disadvantages for Kant’s moral philosophy.<sup>37</sup> Consequently, it is not surprising to find recent efforts to reinterpret Kant’s whole theory of freedom in explicitly compatibilist terms.<sup>38</sup>

---

<sup>36</sup>See *Critique of Pure Reason*, A533/B561, A803/B831; *Critique of Practical Reason*, *Kants gesammelte Schriften*, V, 97, Eng. trans. p. 100.

<sup>37</sup>For an interesting recent formulation of this line of criticism see Terence Irwin, “Morality and Personality: Kant and Green,” in *Self and Nature in Kant’s Philosophy*, edited by Allen W. Wood (Ithaca and London: Cornell University Press, 1984), pp. 31–56.

<sup>38</sup>The most detailed and scholarly of these attempts is by Rolf Meerbote, who interprets Kant in explicitly Davidsonian terms. See his reply to Irwin, “Kant on Freedom and the Rational and Morally Good Will,” *op. cit.*, pp. 57–72; and “Kant on the Nondeterminate Character of Human Actions,” *Kant on Causality, Freedom, and Objectivity*, edited by William A. Harper and Rolf Meerbote (Minneapolis: University of Minnesota Press, 1984), pp. 138–163. Many other contemporary philosophers, most notably Thomas Nagel, present quasi-Kantian accounts of agency while rejecting Kant’s indeterminism.

Although the question of whether Kant's First Critique conception of practical freedom requires indeterminism is quite complex, with texts pointing in both directions, I am inclined to think that it does. I cannot, however, argue the point here.<sup>39</sup> Similarly, I do not intend to discuss either the general issue of indeterminism or the plausibility of the Kantian version. Since we are concerned here only with the implications of transcendental freedom for Kant's moral philosophy, we can set aside these larger issues. For present purposes, the key point is that, even assuming that both practical and transcendental freedom require indeterminism, there remains

---

<sup>39</sup>Although I am no longer happy with all the details of my earlier analyses of this issue in "Practical and Transcendental Freedom in the *Critique of Pure Reason*" and *Kant's Transcendental Idealism*, I still contend that, as far as the interpretation of Kant is concerned, much depends on how one construes Kant's remark in the Canon:

Whether reason is not, in the actions through which it prescribes laws, itself again determined by other influences, and whether that which, in relation to sensuous impulses, is entitled freedom, may not, in relation to higher and more remote operating causes, be nature again, is a question which in the practical field does not concern us (A803/B831).

Taken in connection with Kant's subsequent statement that "we thus know practical freedom to be one of the causes of nature, namely, to be a causality of reason with respect to the will," this suggests that, in the Canon, at least, Kant held that practical freedom would stand even if there were no transcendental freedom. Since transcendental freedom is, by definition, a non-compatibilist or indeterministic conception of freedom, the clear implication is that the detachable conception of practical freedom is not. Thus, one arrives at a compatibilist reading of practical freedom, albeit at the cost of a contradiction between the Canon and the Dialectic. Both the contradiction and the compatibilist reading of practical freedom can be avoided, however, if we keep in mind that the transcendental freedom to which Kant refers in the Canon is construed explicitly as a "causality of reason." As Kant himself states, "transcendental freedom demands the independence of this reason—in respect of its causality, in beginning a series of appearances—from all determining causes of the sensible world" (A803/B831). Denying the "independence" of the causality of reason in this sense would not seem to be equivalent to denying indeterminism; but it would clearly commit one to the view that the will is ineluctably heteronomous. I take Kant's position in *The Critique of Pure Reason* to be an agnosticism with respect to the latter issue, which has nothing directly to do with the determinism-indeterminism question. In order to show that Kant's conception of practical freedom requires indeterminism, it would be necessary to consider his account of "intelligible character," a task that is obviously beyond the scope of this paper.

a significant difference between them. Moreover, this difference is crucial for Kant's moral philosophy.

This becomes clear if we distinguish between independence from determination by any particular desire or inclination and independence from determination by desire or inclination *überhaupt*. Practical freedom involves the first and transcendental freedom the second. Given this distinction, it follows that an agent would be free in the practical (but not in the transcendental) sense if the agent's choices were ultimately governed by some fundamental drive or natural impulse, for example, self-preservation or maximization of pleasure, which can be acted upon in any number of ways but which cannot be contravened. Such an agent would be practically free, possibly even in an incompatibilist sense, because the drive or impulse serves to limit the agent's options rather than to necessitate a given choice.<sup>40</sup> At the same time, however, the agent's choice would be ineluctably heteronomous; since it would be limited to the determination of the best means for the attainment of some end implanted in us by nature. Obviously, such a conception of agency is incompatible with the central tenets of Kant's mature moral philosophy.<sup>41</sup>

The situation with respect to transcendental freedom is quite different. According to this conception, the ground of the selection of a maxim can never be located in an impulse, instinct, or anything "natural"; rather, it must always be sought in a higher order maxim and, therefore, in an act of freedom.<sup>42</sup> Consequently, even if one posits a natural drive such as self-preservation, it remains the case that a transcendently free agent is, *ex hypothesi*, capable of select-

---

<sup>40</sup>The point here is simply that a fundamental drive or impulse is a "standing condition" rather than a cause in the sense of the Second Analogy. The latter, for Kant, is always an event from which another event (the effect) follows *necessarily and in accordance with an absolutely universal rule*" (A91/B124). For my analysis of the Second Analogy, see *Kant's Transcendental Idealism*, Chapter 10.

<sup>41</sup>I argue, however, in "The Concept of Freedom in Kant's Semi-Critical Ethics" (forthcoming in *Archiv für Geschichte der Philosophie*) that it is perfectly compatible with Kant's moral philosophy at the time of the First Edition of the *Critique of Pure Reason*. The key point is that in 1781 Kant had not yet developed his doctrine of autonomy.

<sup>42</sup>Kant develops this doctrine at length in *Religion within the Limits of Reason Alone* in connection with his account of "radical evil."

ing maxims that run directly counter to its dictates. Moreover, since the choices of a transcendently free agent, including those based on desire or inclination, are grounded in a "law" (maxim) which is self-imposed, such an agent would be autonomous in a morally neutral sense. Finally, it should be clear that only an agent that is free in this sense is capable of acting out of "respect for the law," and therefore of acting autonomously in the specifically moral sense on which Kant insists.

## V

We are finally in a position to consider the implications of the presupposition of transcendental freedom for the problem of justification. The basic point is simply that without this presupposition, that is, assuming merely practical freedom, a maxim based on self-interest, happiness, or some such putatively ultimate yet non-moral end or motivational ground could be justified by an appeal to "human nature" or some given determinant of behavior. (The details are irrelevant to the argument.) With it, however, this familiar move is blocked. If self-preservation, self-interest, or happiness is the principle of my behavior, if it dictates my maxims, it is I (not nature in me) that gives it this authority. At least this is the case under the presupposition that I am free in the transcendental sense. Moreover (and this is an essential premise of the entire argument), the justification requirement is still in place. In fact, the presupposition of transcendental freedom not only blocks certain kinds of justification, it also extends this requirement to first principles or fundamental maxims. Since such maxims, like all others are, *ex hypothesi*, freely adopted, it must be possible to offer reasons in support of their adoption. Correlatively, since such principles or maxims are first or fundamental in the sense that they provide the ultimate grounds for the justification of lower order maxims, they obviously cannot be justified by deducing them from some higher order principle.<sup>43</sup>

---

<sup>43</sup>Lewis White Beck, "The Fact of Reason: An Essay on Justification in Ethics," *Studies in the Philosophy of Kant* (Indianapolis: Bobbs-Merrill, 1965), pp. 200–214, provides the best discussion, from a Kantian point of view, of the problems involved in the justification of fundamental practical principles.

How, then, is the rational egoist to deal with this problem? To be sure, the proponent of such a position can continue to assert that it would be reasonable (if not desirable for the egoist) for every rational agent to act according to that principle and, therefore, that it passes the universalizability test in the sense in which the rational egoist acknowledges it. The real question, however, is whether this claim can be justified, given the presupposition of transcendental freedom. Obviously, the claim that it is somehow in one's best interests to act according to the dictates of rational egoism is question begging at best. (At worst it may be simply false.) But the presupposition of transcendental freedom rules out what seems to be the only alternative strategy for justification, namely, the appeal to some given determinant or ultimate fact about human nature, which somehow of itself justifies the adoption of a maxim. Presumably, the same would hold, *mutatis mutandis*, for any other "heteronomous" principle.<sup>44</sup>

Admittedly, the most that this line of reasoning can show is that rational egoism and similar familiar doctrines cannot be rationally justified, at least not if they are combined with the presupposition of transcendental freedom. This is not a trivial result; but it is hardly equivalent to the claim that a rational and transcendently free agent is constrained to acknowledge the validity (as the ultimate norm) of an unconditional practical law. Consequently, even assuming that the moral law, as defined by Kant, is the only conceivable candidate for a practical law, we cannot claim to have established that such an agent is necessarily subject to that law.

Nevertheless, we are finally in possession of the materials needed for such an argument. Although Kant himself never formulated it explicitly, I believe that it is implicit in all of his major writings in moral philosophy. The argument I have in mind is from the assumption of rational and transcendently free agency to the condi-

---

<sup>44</sup>There are obvious affinities between this line of argument and those which appeal to the is-ought distinction and the "naturalistic fallacy." For an important discussion of the relevance of such arguments to Kant's own position see Karl-Heinz Ilting, "Der naturalistische Fehlschluss bei Kant," *Rehabilitation der praktischen Philosophie*, I, edited by Manfred Riedel (Freiburg: Rombach, 1972), pp. 79–97.

tions of the possibility of the justification of the maxims (including the fundamental maxims) of such agents. It proceeds in two stages: the first contends that conformity with an unconditional practical law is a sufficient condition for the justification of these maxims; the second contends that it is also a necessary condition.

The first point I take to be relatively unproblematic. What stronger justification could there be for one's adoption of a maxim than its conformity to an unconditionally valid practical law? If a rule of action is "right" for all rational agents whatever their interests or desires, then, clearly, it is "right" for me. Again, if my reason for X-ing is that it is dictated by such a law (in Kant's deontic terms, that it is my duty), then I have all the justification I would conceivably need for X-ing. This is not to deny that there may be grave difficulties determining exactly what such a law requires in a given instance (what my duty is), and, therefore, that Kant's moral philosophy may run into severe difficulties in this regard. The present point is only that *if* a maxim can be shown to meet this requirement then that maxim has been fully justified.

Obviously, the main difficulties concern the claim that this requirement is a necessary condition for justification. In dealing with this issue, it will be helpful to begin with the consideration of a familiar yet misguided criticism of Kant's moral theory. Couched in terms of the present discussion, the claim is that the requirement (at least as here construed) is too strong. If, so the argument goes, the only legitimate reason for adopting a maxim were its conformity to a practical law binding upon all rational agents, regardless of their interests or desires, then it would seem that no maxim to pursue one's interests or desires could ever be justified. But this is patently absurd. Thus, even if it be granted that conformity to a practical law is a sufficient condition for the justification of one's maxims, it is certainly not also a necessary condition. To claim otherwise is to commit oneself to the doctrine that only actions performed "for the sake of duty" are justifiable; and this is to conflate justifiability with moral worth.

Although there is undoubtedly a strand in Kant's moral philosophy that suggests this line of interpretation and criticism, it does not reflect his considered opinion. What this reading neglects is the centrality for Kant of the distinction between the permissible and

the obligatory.<sup>45</sup> Not surprisingly, then, it also fails to recognize that the moral law is intended as a criterion of the former as well as of the latter.<sup>46</sup> This is, of course, a large topic in its own right. Indeed, it calls to mind all of the familiar difficulties concerning Kant's distinctions between positive and negative, imperfect and perfect duties. I introduce it here only because it indicates that, rather than ruling out as illegitimate all desire or interest based maxims, the notion of conformity to a practical law is intended by Kant as a criterion for determining which maxims of this (or any) sort are permissible. Moreover, since it seems clear that no maxim could be regarded as justified if it were not at least shown to be permissible, it follows that establishing this claim is equivalent to showing that conformity to a practical law functions as a necessary condition for the justification of maxims.

Permissibility, like other deontic notions, has both a specifically moral and a morally neutral sense. In the former case it encompasses whatever is not contrary to duty and in the latter whatever is allowable within a given context or in light of some pre-given end (in accordance with the "rules of the game").<sup>47</sup> Presently, however, we are concerned merely with the conditions of the justification of the desire or interest based maxims of transcendently free rational agents, that is, agents for whom the choice of such maxims both requires rational justification and is attributed to an act of freedom. Given these constraints, it is apparent that permissibility cannot be construed as a function of desires or interests, even the most fundamental ones. In other words, we are not looking for a rule or set of rules which determine what is permissible within the framework of some presupposed end. On the contrary, what must be determined is the rule or set of rules governing the pursuit of any end at all, including desire or interest based ends. In view of the "transcendental" function of such a rule or set of rules (its function with respect to end setting *überhaupt*), it is also apparent

---

<sup>45</sup>See *Groundwork*, *Kants gesammelte Schriften*, IV 438, and the *Metaphysics of Morals*, VI, 422.

<sup>46</sup>For a discussion of the issue see Paton, *The Categorical Imperative*, pp. 141–142, and Beck, *A Commentary*, p. 122.

<sup>47</sup>Kant himself suggests such a distinction in the *Critique of Practical Reason*, *Kants gesammelte Schriften*, V. 11n.



that it must be both universal and “formal” in the specifically Kantian sense. That is, it must not only apply to all transcendently free rational agents, it must also apply to them regardless of what desires or interests they may happen to have. But such a rule or set of rules is precisely what Kant understands by a practical law. Consequently, it must either be denied that the maxims of transcendently free agents can be justified at all (which amounts to a denial of rationality) or it must be acknowledged that conformity to practical law is the criterion governing the selection of the maxims of such agents. Combining this result with the further claim that the moral law is the required principle, we have the Kantian argument for the “analytic” connection between transcendental freedom and the moral law.

Since the above analysis turns largely on the distinction between the rationality and the freedom of an agent, it might itself seem problematic as Kantian exegesis. Such a conclusion, however, would be erroneous. Although Kant only makes this distinction fully explicit in *Religion within the Limits of Reason Alone*,<sup>48</sup> it is implicit in his thought from the *Groundwork* on.<sup>49</sup> Admittedly, only a rational agent can be free in the transcendental (or even the practical) sense, but an agent is not free in the transcendental sense simply in virtue of being rational. At least freedom in this sense cannot be derived from rationality; and, therefore, neither can the validity of the moral law. Unfortunately, the whole point is usually missed by Kant's critics. Starting with the reasonable assumption that a Kantian justification of morality must somehow demonstrate the irrationality of rejecting the categorical imperative, these critics tend to assume that the argument to this end must proceed simply from the

---

<sup>48</sup>Kant's clearest statement on this point occurs in connection with his account of the distinction between the predisposition (*Anlage*) to *humanity* in man, “taken as a living and at the same time a *rational* being” and the predisposition to *personality* in man, “taken as a rational and at the same time accountable being,” *Kants gesammelte Schriften*, VI, 26–28.

<sup>49</sup>I take this distinction to be implicit in the distinction which Kant suggests in the *Groundwork* between rational beings *simpliciter* and rational beings possessed of a will. *Kants gesammelte Schriften*, IV, 448–459. The same distinction is also operative in the *Critique of Practical Reason*, *Kants gesammelte Schriften*, V, 32. For a discussion of the significance of this distinction see Dieter Henrich, “Die Deduktion des Sittengesetzes,” esp. pp. 91–100.

concept of a rational being. This is not the case, but it is only by focusing explicitly on the Reciprocity Thesis that it becomes clear why it is not.

## VI

Given the preceding analysis, we can now deal with the common objection that the Reciprocity Thesis, particularly as presented in the *Groundwork*, leads Kant to the devastating consequence that we are free only insofar as we act in obedience to the categorical imperative. This consequence is devastating not only because it entails that we are not responsible for either our immoral or our morally neutral actions, but also because it suggests that even our morally good actions (actions performed for the sake of duty) are due *ultimately to a fortuitous lack of interference by nature* (in the guise of sensuous inclination) with the autonomous workings of pure practical reason. After all, if a free will is defined as one governed by the moral law, and if, as Kant suggests, this is analogous to the way in which natural phenomena are governed by the laws of nature, then it would seem that a free will could no more violate the moral law than a falling body could violate the law of gravity. Correlatively, if a non-moral or heteronomous will is subject to the laws of nature, then there is no way to understand how a will that is not already moral could choose to become such. In short, by distinguishing so strongly between nature and freedom and by defining the freedom of the will in terms of its subjection to the moral law, Kant seems to have made it impossible to provide a coherent account of either immoral or moral action.<sup>50</sup>

The standard strategy for defending Kant against this line of objection is to admit that it applies to the *Groundwork* and to deny that it applies to the later treatment in *Religion within the Limits of Reason Alone*, where Kant offers an account of moral evil in light of

---

<sup>50</sup>The most detailed and powerful formulation of this line of criticism in the recent literature is by Prauss, *Kant über Freiheit als Autonomie*, esp. pp. 60–115. Versions of it are found, however, in a large number of commentators. Indeed, as Prauss points out, it can be traced back to Kant's own contemporaries.

the *Wille-Willkür* distinction.<sup>51</sup> The basic idea is that *Willkür*, as spontaneous, is free either to determine itself to act in accordance with the dictates of *Wille* (the stern call of duty) or to subordinate these dictates to the demands of inclination. The claim, in other words, is that what the *Groundwork* presents as heteronomy and opposes to autonomy is seen in *Religion within the Limits of Reason Alone* as itself an expression of freedom.

There can be little doubt that the account of freedom in *Religion within the Limits of Reason Alone* has a subtlety and depth that are lacking in both the *Groundwork* and the *Critique of Practical Reason*. We have already seen, however, that rather than precluding this conception of freedom, the Reciprocity Thesis, as formulated in these earlier works, requires it in the sense that it is only by construing freedom in this way that the argument can be made to work. Admittedly, this does not of itself prove that Kant actually held such a conception at that time, but it certainly suggests that he could have, and it puts the burden of proof on those who would deny it.

There appear to be three aspects of Kant's account in the *Groundwork* to which the critic can appeal in support of this denial. Since all three of them have already been noted, we need only recall them here. First and foremost is the language of the Reciprocity Thesis itself. By explicitly identifying a free will with a will under the moral law, Kant certainly seems to leave no room for any free action that does not conform to the law. Second, this impression is greatly reinforced by the unfortunate analogy between the moral law and a law of nature. Finally, there is the apparent identification of heteronomous action and action in accordance with (or governed by) the laws of nature. This identification suggests that a heteronomous will can be neither free nor morally responsible, and this, in turn, raises the perplexing question of how such a will could ever become either free or morally good. Although not free (because heteronomous) is it free to become free?

In dealing with the first two aspects, it is obviously crucial to

---

<sup>51</sup>See Silber, "The Ethical Significance of Kant's Religion," LXXXV, CXXVII–CXXVIII.

determine with some precision the sense in which a free will is supposed to be subject to, or governed by, the moral law. As we have already seen, in the *Critique of Practical Reason* Kant maintains that only a formal principle (later identified with the moral law) is “competent to determine it (a free will) necessarily.” This locution once again suggests the very problem currently under consideration. Being determined necessarily by the moral law seems to mean being subject to it in precisely the same sense in which a physical object is subject to the laws of nature. And this, of course, rules out the possibility of any deviation from the law. It is, however, not only unnecessary, it is also implausible to take Kant to be making any such claim. Since “determine necessarily” means simply to possess a lawlike status for a free will (to be “objectively necessary”), all that Kant is claiming here is that only a formal principle (the moral law) can serve in this capacity for a free will.

The question thus becomes what is involved in serving in this capacity and the answer is quite apparent. It can mean only that the law provides a norm or standard in terms of which the choices of a free will are justified before the bar of reason. Material practical principles cannot do the job because they presuppose an object of desire as the determining ground of the will and, as we have seen, a free will is (by definition) responsible for the selection of any such objects as its ends. It hardly follows, however, from the fact that the moral law is the norm or standard for a free will that such a will is not “capable” of failing to live up to this norm. As Kant frequently insists, although “objectively necessary,” the moral law is nonetheless “subjectively contingent.”<sup>52</sup> Consequently, we are free to act heteronomously, to make the satisfaction of our desires the basis of our choice. In so doing we are, at least according to Kant’s moral theory, misusing our freedom; indeed, we are misusing ourselves in that we are treating our “higher” or “proper self” [*das eigentliche Selbst*] merely as a means for the satisfaction of our “lower” or sensuous nature. Nevertheless, this misuse of freedom is still very much an act of freedom, and there is nothing in Kant’s theory that requires us to think otherwise.

Perhaps the single most important Kantian text bearing on this

---

<sup>52</sup>See *Kants gesammelte Schriften*, IV, 413–414.

issue is the discussion in the Introduction to the *Metaphysics of Morals*. Kant there first defines freedom (the positive concept) as “the power of pure reason to be of itself practical.”<sup>53</sup> Then, later, after introducing the *Wille*—*Willkür* distinction, and claiming that *Wille* cannot be properly regarded as either free or unfree since it deals with legislation rather than action, he writes

Freedom of choice, however, cannot be defined as the capacity for making a choice to act for or against the law (*libertas indifferentiae*), as some people have tried to define it, even though choice as a phenomenon gives frequent instances of this in experience. For freedom (as it first becomes known to us through the moral law) is known only as a negative property within us, namely, the property of not being constrained to action by any sensible determining grounds. . . . But we can see clearly that although experience tells us that man as a sensible being exhibits a capacity to choose not only in accordance with the law but also in opposition to it, yet his freedom as an intelligible being cannot be thus defined, since appearances can never enable us to comprehend any supersensible object (such as free choice is). . . . For it is one thing to admit a tenet (of experience) and quite another to make it both the defining principle (of the concept of free choice) and the universal mark distinguishing free choice from *arbitrio bruto s. servo*, since in the first case we do not assert that the mark necessarily belongs to the concept, which we must do in the latter case. Only freedom in relation to the internal legislation of reason is properly a capacity; the possibility of deviating from it is an incapacity. How, then, can the former be explained by the latter?<sup>54</sup>

Already Reinhold had objected to this formulation by presenting a dilemma. According to Reinhold's analysis, if the only concept of freedom derivable from the moral law is that of the self-activity (*Selbsttätigkeit*) of reason, then the presumed “capacity” to act immorally is not only an incapacity but an impossibility. If, on the other hand, freedom is construed as the capacity of the *person* for self-determination, then the “capacity” to act immorally is not a mere incapacity but rather the very same capacity without which

---

<sup>53</sup>*Kants gesammelte Schriften*, VI, 214, Eng. trans. *The Metaphysical Principles of Virtue*, translated by James Ellington (Indianapolis, New York: Bobbs-Merrill), 1964, p. 12.

<sup>54</sup>*Ibid.*, p. 226. Eng. trans. p. 26.

moral action cannot be thought.<sup>55</sup> Otherwise expressed, Reinhold's complaint seems to be that the proper concept of freedom must be a morally neutral one, and that this is incompatible with Kant's insistence that our understanding of freedom is derived entirely from our consciousness of the moral law. Quite recently, Gerold Prauss has raised similar objections. Prauss, however, also claims that Kant's account of freedom in the passage currently before us marks a regression from the standpoint of *Religion within the Limits of Reason Alone*, where he at least attempted (albeit unsuccessfully) to provide an account of immoral action in terms of freedom, back to that of the *Groundwork*, where such action is seen as a product of the heteronomy of nature.<sup>56</sup>

We can readily accept Prauss's assertion of the agreement of the account of freedom in this passage with that of the *Groundwork*, although not his characterization of it as a "regression," and certainly not his analysis of its implications. The key term in Kant's account is obviously 'power' or 'capacity' (*Vermögen*). By the "power of reason to be of itself practical" Kant means first of all its capacity to provide a binding law for the will.<sup>57</sup> Reinhold is correct in suggesting that if this were all that Kant means by freedom, then the freedom to disobey the law has not been established. There is, however, no need to accept this result nor, therefore, the terms of Reinhold's dilemma. Since ought implies can (at least for Kant), the capacity of pure reason to be practical, that is, to provide a binding law, entails the capacity of a free agent to obey the dictates of this law. This is precisely the point that Kant makes when he remarks that through the moral law we are aware of freedom as a "negative property . . . of not being constrained to action by any sensible de-

---

<sup>55</sup>Karl Leonhard Reinhold, "Einige Bemerkungen über die in der Einleitung zu den 'Metaphysischen Anfangsgründen der Rechtslehre' von I. Kant aufgestellten Begriffe von der Freiheit des Willens," in *Materialien zu Kants "Kritik der praktischen Vernunft"*, edited by Rüdiger Bittner and Konrad Cramer (Frankfurt am Main: Suhrkamp Verlag, 1975), pp. 310–323.

<sup>56</sup>Prauss, *Kant über Freiheit als Autonomie*, p. 112. Prauss also offers a critique of Reinhold's own account, pp. 84–92.

<sup>57</sup>For a similar analysis of this text and its relevance to the general problem of the possibility of free and yet non-moral action for Kant see Nelson Potter, Jr., "Does Kant have Two Concepts of Freedom?" *Akten des 4. Internationalen Kant-Kongresses*, 590–596.

termining grounds." Thus, freedom is understood as the will's capacity to follow its own self-imposed laws, which requires an independence from constraint by any sensible determining grounds. In the preferred jargon of Kant's later moral philosophy, freedom is construed as the capacity of *Willkür* to obey the dictates of *Wille*. Once again, it should be obvious that the possession of such a capacity is perfectly compatible with the failure to actualize it. Kant acknowledges this, but he also insists that this "capacity" to fail is really an incapacity (presumably because it cannot be ascribed to perfectly rational beings) and, therefore, should not be regarded as definitional of freedom. Reinhold and Prauss to the contrary, this does not at all entail that such failure is not itself an expression of freedom.

The third and final aspect of the problem concerns Kant's tendency, particularly in the *Groundwork*, to equate heteronomy with subjection to the laws of nature. In response to this it should suffice to note that there is no need to take the claim that the heteronomous will is "subject" to the laws of nature to mean anything more than that the inclinations and desires upon which it bases its choice are themselves products of nature. Subjection to the laws of nature in this sense is perfectly compatible with the Kantian conception of freedom. It does not follow from the fact that the inclinations and desires on the basis of which one chooses to act are products of nature, that the act of choice itself, through which they are "incorporated" into the maxim of the will, is likewise such a product.

## VII

Even if sound, the argument offered here for the Reciprocity Thesis hardly suffices to establish the Kantian version of morality. It shows only that we cannot both affirm our freedom (construed in the transcendental sense) and reject the categorical imperative. In this respect it can be said to have established the price of moral skepticism. The problem, of course, is that this price (the rejection of transcendental freedom) is one that the moral skeptic (or the rational egoist) is more than willing to pay. This is particularly true in view of the notorious difficulties in Kant's attempt to reconcile this freedom with the universal sway of the principle of causality.

Nevertheless, the question of transcendental freedom must at least be faced by anyone who wishes to criticize the Kantian attempt at the justification of morality. This attempt cannot be dismissed in the casual manner of Philippa Foot, who contends that there is nothing irrational or inconsistent in the rejection of the categorical imperative, or, at least, that no one has ever shown that there is,<sup>58</sup> while also insisting elsewhere that “a reason for acting must relate the action directly or indirectly to something the agent wants or which it is in his interest to have. . . .”<sup>59</sup> Foot is certainly consistent here, but in a way that sidesteps the main thrust of Kant’s position. Given her conception of agency, it not only would not be irrational to reject the categorical imperative, it would be metaphysically impossible to obey it. Perhaps it is, but simply to assume this is the case, is to beg the whole question.

In the last analysis, then, Kant’s moral theory stands or falls with the metaphysical doctrine of transcendental freedom. As the Reciprocity Thesis makes clear, this freedom is not only a necessary, it is also a sufficient condition of morality as Kant conceives it. Consequently, if this freedom be denied, nothing remains save a rather complex and convoluted analysis of the presuppositions of a set of illusory beliefs. If, on the other hand, it be granted, then the validity of the moral law follows. This same reciprocity, however, suggests that it might very well be impossible to establish either one without presupposing the other, which would mean that Kant’s attempt to justify morality is caught in a vicious circle from which there is no escape. Kant himself raises the spectre of just such a circle in the *Groundwork*, but claims to be able to avoid it. Whether

---

<sup>58</sup>This is the central thesis of her influential paper, “Morality as a System of Hypothetical Imperatives,” *The Philosophical Review* 81 (1972), pp. 305–316. Reprinted in *Virtue and Vices and Other Essays in Moral Philosophy* (Berkeley and Los Angeles: University of California Press, 1978), pp. 157–178.

<sup>59</sup>“A Reply to Professor Frankena,” *Virtues and Vices*, p. 179. It should perhaps also be noted here that Foot defends an essentially Humean theory of agency, including a version of compatibilism in several of the essays included in this volume. This is not, of course, to say that she is wrong in this regard, but only that the question of the cogency of Kant’s analysis of morality is inseparable from the question of the cogency of his theory of agency.



he is successful, either there or in his fresh treatment of the problem in the *Critique of Practical Reason*, is a larger issue, which cannot be dealt with here.

*University of California, San Diego*