The Idea of Freedom and Moral Cognition in Groundwork *III*

Sergio Tenenbaum

1 Introduction

Although the relation between freedom and the moral law is central to Kant's moral philosophy, it is often difficult to explain precisely the nature of this relation in Kant's work, and how Kant's thought evolved in this matter from his pre-Critical writings to his later work. All commentators agree that at least in all his Critical works, Kant endorses some version of what Henry Allison calls "The Reciprocity Thesis", the thesis that freedom and the moral law imply each other. However, there's significant controversy on how various arguments in Kant's corpus are supposed to move us from the fact that we are free to the fact that we are bound by the moral law, or vice-versa. Particularly puzzling is what seems to be a major shift in Kant's position on this relation. It seems that in various works up to, but not including, the *Critique of Practical Reason*,[1] Kant seems to think that he has independent grounds to establish that we're free and that he can use this fact as some kind of foundation for the moral law.[2] However, there could be little doubt that Kant later came to deny that we have any access to the fact that we are free independently of the moral law. In the *Critique of Practical Reason*, Kant says that

> the moral law is the *ratio cognoscendi* of freedom. For, had not the moral law *already* been distinctly thought in our reason, we

[1] For an account of the evolution of Kant's thoughts on this issue from the pre-critical period on, see Dieter Henrich, "The Concept of Moral Insight and Kant's Doctrine of the Fact of Reason" in his *The Unity of Reason*, Harvard University Press, 1994.

[2] For very different views of the argument that accept some version of this claim, see Allen Wood, *Kant's Ethical Theory* (New York: Cambridge University Press, 1999), ch. 5, Korsgaard, "Morality as Freedom" in her *Creating the Kingdom of Ends* (New York, Cambridge University Press), David Sussman, "From Deduction to Deed: Kant's Grounding of the Moral Law" *Kantian Review* **30**: 52-81, and Henry Allison, *Kant's Theory of Freedom* (New York: Cambridge University Press, 1990).

> should never consider ourselves justified in *assuming* such a thing as freedom. (KpV 4n, emphases in original)[3]

And in the *Religion*, Kant seems to claim that there is simply *no* form of reasoning that could lead someone to infer that they are bound by the moral law, and again, that our awareness of the moral law is the source of our awareness of our freedom, rather than the other way around:

> Were this [moral] law not given to us from within, no amount of subtle reasoning on our part would produce it or win our power of choice over to it. Yet this law is the only law that makes us conscious of the independence of our power of choice (*Willkür*) from determination by all other incentives (or our freedom).

Of course one could read claims such as "the moral law makes us conscious of our freedom" as metaphysical claims, but given that Kant clearly thinks that freedom is the *ratio essendi* of the moral law, we must read these assertions as making epistemological claims.

On the other hand, in the *Groundwork* Kant seems to proceed in the opposite manner. Many, if not most, of the commentators take these appearances at face value and come to the conclusion that Kant's thought

[3]References to Kant's works are to the appropriate volume of *Kants gesammelte Schriften, herausgegeben von der Deutschen* (formerly *Königlichen Preussischen*) *Akademie der Wissenschaften* (Berlin: Walter de Gruyter (and predecessors), 1902), with the exception of the Critique of Pure Reason and Lectures on Ethics. References to the *Critique of Pure Reason* are to the standard A and B pagination of the first and second edition. Reference to the *Lectures on Ethics* is to *Eine Vorlesung ber Ethik*, edited by Gerd Gerhardt (Frankfurt am Main: Fischer Verlag, 1990). Specific works are cited by means of the abbreviations below. I have used the English translations mentioned below with occasional minor changes. I have provided the page number of the German edition and the English translation (the latter in parentheses) whenever the latter did not include the German pagination in the margins.

G	*Groundwork of The Metaphysics of Morals*, trans. Mary Gregor (Cambridge: Cambridge University Press, 1998).
KpV	*Critique of Practical Reason*, trans. Mary Gregor (Cambridge: Cambridge University Press, 1997).
KrV	*Critique of Pure Reason*, trans. Paul Guyer and Allen Wood (Cambridge: Cambridge University Press, 1999).
MS	*Metaphysics of Morals*, trans. Mary Gregor (Cambridge: Cambridge University Press, 1996).
R	*Religion within the Boundaries of Mere Reason*, trans. Allen Wood and George di Giovanni (Cambridge: Cambridge University Press, 1998).
VE	*Lectures on Ethics*, trans. Louis Infield (Indianapolis: Hackett Publishing Co., 1981).

underwent a major reversal between the *Groundwork* and the *Critique of Practical Reason*; that is, between 1785 and 1788. However, unlike his famous repudiation of some of his pre-Critical views, Kant seems to be unaware, or at least fails to acknowledge, that his view has been radically transformed. How could Kant have been oblivious to such a major change in his understanding of the relation between freedom and the moral law? Had he not noticed it himself? Or did he think that this was not an important enough change to be worth spilling some ink over? No doubt these kinds of considerations are not decisive; we might think that Kant is not being very forthright about his view, or that he expects his readers to note changes on their own. But it is worth noting that there is not much else that changes in relation to these issues, or at least not much that could be relevant for the argument one way or the other. Whatever reasons Kant had to change his mind in 1788, they were already available to him in 1785; no explicit new views about freedom, morality, or the relation of the sensible and intelligible world are introduced that could justify such a shift. At any rate, it seems safe to say that there is much to be said for trying to understand the argument of *Groundwork III* in such a way that it does not conflict with the later views in the *Critique of Practical Reason*. And this is particularly true, if, as I argue below, the arguments commonly attributed to Kant that supposedly find a route from a cognition of freedom to a cognition of the moral law are neither philosophically appealing nor compatible with much that Kant holds dear.

My aim in this paper is relatively modest. I will limit myself mostly to trying to give an account of Kant's first steps in the argument and, in particular, his understanding of what it is to act under the idea of freedom, and how we arrive at, and what follows from, the conclusion that an agent must act under the idea of freedom. I look at a couple of interpretations of these opening passages of *Groundwork III* that, if correct, would pave the way for an interpretation of *Groundwork III* as containing an argument that moves from the fact that we are free to a proof of the moral law. I hope to show that these interpretations fail both on textual and on philosophical grounds. I then try to present an alternative interpretation of what Kant means by "acting under the idea of freedom" as well the use to which he tries to put this notion. This interpretation, I argue, gives us no reason to think that Kant was using this notion in *Groundwork III* to set up this kind of argument from freedom to the moral law. I then briefly sketch how we can read the rest of *Groundwork III* in a way that is entirely (or at least mostly) compatible with the *Critique of Practical Reason*. If I am right, at least as early as the *Groundwork*, Kant no longer held that we have any

access to the fact that we are free other than via our awareness of the moral law.[4] If this is true, Kant should be seen as an unlikely ally for anyone who wants to derive our commitment to the moral law from a conception of rational agency that does not already presuppose a commitment to morality. This, I hope, does not show that Kant's moral philosophy is uninteresting or even unambitious with regard to what it tries to establish regarding the relation between freedom and the moral law. But what's most interesting and controversial about Kant's views in this area is the reciprocity thesis itself, not any views about how the reciprocity thesis can provide us access to the moral law.

2 Korsgaard on "Under the Idea of Freedom"

Kant's claim that we must act under the idea of freedom is so well-known that it hardly needs citing. But here it is, once again:

> I say now: every being that cannot act otherwise than *under the idea of freedom* is just because of that really free in a practical respect; that is, all laws that are inseparably bound up with freedom hold for him just as if his will had been validly pronounced free in itself and in theoretical philosophy. (G 448)

In "Morality as Freedom", Christine Korsgaard advances a deservedly influential reading of this passage. According to Korsgaard, Kant is pointing out that even if we were to learn that determinism is true, when deciding what to do, it would make no difference to our deliberations. Even if we were to learn that determinism is true, we would still have to deliberate as if we were free. Here is how Korsgaard puts the claim:

> The point is not that you must *believe* that you are free, but that you must choose *as if* you were free. It is important to see that this is quite consistent with believing yourself to be fully determined... Kant's point, then, is not about a theoretical assumption necessary to decision, but about a fundamental feature of the standpoint from which decisions are made. It follows from

[4] Of course, I am not the first interpreter to suggest that Kant's position in the *Groundwork III* is consistent with his views in the *Critique of Practical Reason* in this manner. For an important precedent, see Paton, H. *The Categorical Imperative* (London, Hutchinson & Co., 1958).

> this feature that we must regard our decisions as springing ultimately from the principles that we have chosen. We must regard ourselves as having free will.[5]

It is worth noting a couple of things immediately about Korsgaard's interpretation. First, whether or not Korsgaard is right in claiming that "acting under the idea of freedom" is "consistent with believing yourself to be fully determined", Kant is clearly not in general so cavalier about the implications that the truth of determinism might have for the validity of moral law. In particular, at least in some sense of "fully determined",[6] it is not true that a being who is fully determined in that sense is a being such that "all laws that are inseparably bound up with freedom hold for him". And this is certainly Kant's view before 1785. Just to give an example, this is what Kant presents as a disadvantage for the empiricist (the antithesis) side of the antinomies; that is, the side that accepts that there is no freedom:

> There is *first* no such practical interest from pure principles of reason as morality and religion carry with them. Mere empiricism seems rather to take all the power and influence away from both. For ... if our will is not free ... then the *moral* ideas and principles also lose all validity, and they collapse along with the *transcendental* ideas that constitute their theoretical support. (KrV, A468/B 496)

Kant claims here that those who do believe in a rampant determinism as a conclusion of the argument of the antithesis must deny that moral principles have any validity. There Kant does not seem to think that they could appeal to the fact that we must act under the Idea of freedom and have a peaceful coexistence between theoretical empiricism and the practical interest of reason. Moreover he seems to advance a very similar claim in *Groundwork III* itself:

> Philosophy must therefore assume that no true contradiction will be found between freedom and natural necessity in the very same human actions, for it cannot give up the concept of nature any more than that of freedom ... [I]f even the thought of freedom

[5] "Morality as Freedom": 162-3.

[6] Needless to say, the issue is more complicated than I can do justice to here. But if "fully determined" is supposed to rule out the possibility of any determination of my actions other than its empirical determination in accordance to the laws of nature, then the full determination of my action is incompatible with my being bound by the moral law.

> contradicts itself or contradicts nature which is equally necessary, it would have to be given up altogether in favour of natural necessity. (G 456)

This is very far from the claim that we can act under the idea of freedom independently of what we actually believe about freedom and determinism.[7]These passages in themselves should cast some suspicion on the claim that Kant endorses the argument that Korsgaard ascribes to him. After all, Kant does think that one of the major achievements of the metaphysics of transcendental idealism is to make room for morality.

One might try to avoid these difficulties by suggesting that we moderate Korsgaard's claim that the truth of determinism is irrelevant for the possibility of acting under the idea of freedom; we could advance a similar view on her behalf by arguing instead that *as long as we at least remain agnostic about the truth of determinism*, we must choose as if we are free. So perhaps a slightly modified verson of Korsgaard's interpretation could accommodate Kant's views about the relation between the truth of determinism and free agency. Now I'm not sure that Korsgaard's interpretation can be reconstructed in this manner; after all, her claim is that the nature of the practical standpoint is such that in deciding you must *thereby* regard yourself as free. It is unclear how any theoretical belief could alter this essential relation between deciding and regarding oneself as free. But whether or not one can revise Korsgaard's interpretation in this manner, I think there are more important reasons to cast doubt on the textual basis of her interpretation, but these problems will be clearer when we look at whether her interpretation of Kant's argument is philosophically persuasive.

So let us look more closely at how Korsgaard reconstructs this part of Kant's argument. We can separate two steps in the argument as Korsgaard interprets it. First, she tries to establish that we must act under the idea of freedom understood as explained above; that is, the first step of the argument should show that when we act we must assume that we are free

[7]The label "determinism" is somewhat confusing in the context of Kant's view; as Allen Wood has famously pointed out, Kant seems to want to show "the compatibility of compatibilism and incompatibilism" ("Kant's Compatibilism" in '*Self and Nature in Kant's Philosophy*, edited by Allen Wood (Cornell: Cornell University, 1984): 74). Kant does think that freedom is compatible with determination by laws; determination by rational laws is free agency *par excellence*. What Kant thinks is incompatible with freedom is being determined by natural laws alone; that is, freedom is not compatible with our action being determined by nothing other than natural necessity. When I use "determinism" throughout the paper I mean to refer to the view that all our actions are fully determined by the laws of nature and nothing else.

even if we believe (or cannot rule out the possibility), from a theoretical standpoint, that determinism is true. Secondly she tries to show that the moral principles hold for those who must act under the idea of freedom, on this interpretation of "under the idea of freedom".

Let us start with the first step. The crux of this part of Korsgaard's argument[8] is an example that supposedly shows how determinism is irrelevant to our deliberations. Korsgaard imagines that our minds are being controlled by some kind of device in our brain programmed by some neuroscientists. She says

> (Y)ou get up and and decide to spend the morning working. You no sooner make the decision than it occurs to you that it must

> world is not the best of all possible worlds. But efforts to guess how the actual world is different from the world that an infinitely wise God would have created cannot help you decide what to do. In order to do anything you must simply ignore the fact that the universe is a godless place and decide what to do— just in the same way as you would decide if the universe were created by an infinitely wise God.

This argument is certainly not an improvement over Kant's argument for belief in God. There is indeed something that the argument gets right: the possibility that there's no infinitely wise God is something that ought to be taken into account in virtually none of my decisions. It's also quite clear what the argument gets wrong; whether or not there is an infinitely wise God ought to be *irrelevant* to nearly all my decisions. I do not act under the Idea of divine wisdom or the Idea of a purposeless universe; I just find my beliefs on these issues to have no bearing on most practical questions. One can't argue from the irrelevance of a certain consideration to our deliberations to the claim that we ought to act as if the consideration were false.

How far can we extend this point to Korsgaard's original argument? At first blush, it seems that Korsgaard is also moving from the irrelevance of a certain consideration to the conclusion that we ought to act as if this consideration were false. After all, what we learn in considering the case that we know that we have been programmed to act in a certain way is that this kind of information ends up having no bearing on what to decide. The truth or falsity of determinism is typically not a reason for or against spending the morning working. So if you are deliberating well (if the programmers didn't make your deliberation malfunction in this particular way), you will not take into account the fact that your actions are predetermined. But, equally, if you were to learn that you were free, if *per impossibile* your freedom were to be proven theoretically, that would also make no difference in your deliberations on whether you should spend the morning working. For, in the normal course of events, the fact that you are free is irrelevant to the question of whether you should work, and thus, one could equally say that we must act under the Idea of determinism–that when deliberating, you must deliberate as if your actions were fully determined.

There is, however, an obviously important difference between the two arguments. Determinism supposedly tells us something about whether I have real alternatives or not, and awareness of the fact that there are no real alternatives does seem to be relevant to my actions. If I realize that it is not in my power to win the New York marathon, it will not be a live option

for me in my deliberations. And it might be tempting to assume that the realization that determinism is true will be a kind of devastating extension of the realization that I cannot win the New York Marathon. The issue is not whether the truth or falsity of determinism should figure in the content of one's deliberation, but whether the fact, or the possibility, that there are no real alternatives in this sense should in any way be relevant to your choices. And the point that Korsgaard seems to be making is that the answer is "no". When you deliberate, you must assume that these alternatives are really open to you, even if you are aware that they might not really be. And, in that sense, you are deliberating as if you were free; you must assume that all these alternatives are really open to you even if you know or suspect that they are not. So on this reading of Kant's argument "acting under the idea of freedom" would amount to something like "acting as if the alternatives are really open", or perhaps "acting as if one's choice of maxims were a genuine one".

Now let us start with a rarely noticed problem with this understanding of what it is to act under the idea of freedom. As we noted above, showing that rational agents must act under the Idea of Freedom is the first step in some kind of argument for the validity of the categorical imperative, or the applicability of the moral law to us. At the very least it should follow, with the aid of the Reciprocity Thesis, that a being who acts under the Idea of Freedom is such that "all laws that are inseparably bound up with freedom hold for him". But how would the second step proceed if we accept Korsgaard's reading of "acting under the idea of freedom"? Now there are two basic ways to understand how Korsgaard is reading the claim that we *must* act under the idea of freedom. On the one hand, we could take "must" to denote either some kind of psychological inescapability or some kind of conceptual necessity. On the other hand, "must" could denote a normative demand, a claim that we *ought* to behave as if we are free, or that somehow in acting we undertake a commitment to act as if we are free.

Let us begin with the non-normative understanding, which seems to be the more natural one. As I said above, on the non-normative side, we could think of the "must" as denoting either conceptual necessity or psychological inescapability. I'll focus on the possibility that we are talking here of a conceptual connection; indeed, Korsgaard's argument seems to postulate a conceptual connection between conceiving of myself as settling between alternatives, and conceiving that any of these alternatives could have been brought about through my agency. However, the problem I will raise can only get worse by making the relation one of psychological inescapability.

How do we show that if we must act under the idea of freedom, we must be bound by the moral law? The obvious answer seems to be that all that we need to do now is to exploit one direction of the Reciprocity Thesis, the claim that free agents must be bound by the moral law. We could now show that all rational agents, including human agents, are bound by the moral law, since they must act just as if they were free. However, the argument assumes that if **q** follows from **p**, and that, in certain contexts, I must act as if **p** were true, then it follows that I must act as if **q** were true too in these contexts. Since the relation "x must act as if **p**" is a bit obscure,[11] it's worth noting that this inference does not work even for "x believes that **p**", since belief is not closed under logical entailment. That is, even if we show that it is a matter of conceptual necessity that one believes that **p**, it will not thereby follow that is a matter of conceptual necessity that one believes a particular consequence of **p**, since it is not a matter of any kind of necessity that one believes all the logical implications of one's beliefs.

Of course, one might argue that there is some kind of normative demand that one accept the implications of what one believes. But could there really be such a normative demand grounded *solely* on the fact that this is an implication of a proposition that one believes (or a proposition that one must act as if one believed in it)? The fact that one believes a proposition can't be by itself a reason to believe any of its implications. Otherwise, given that any belief entails itself, the mere fact that I believe something would give me at least *some* reason to believe it; we could thus rest assured that none of our beliefs are groundless. But if the mere fact that I believe **p** can't constitute a reason to accept what follows from **p**, the mere fact of my having to act as if I believe that **p**, can't give me a reason to accept what follows from **p**. The demand must ultimately rest on the fact that the grounds that justify one's beliefs in a certain proposition will *a fortiori* justify the consequences of this proposition. But in this case no similar relation holds between the grounds in which we act as if we are free and the grounds to accept the moral law. After all, on this account we do not accept that we are free on any grounds; it is simply an inevitable consequence of

[11] This is sometimes understood as "believing from a practical standpoint". For more specific doubts about whether this notion can do the job that is supposed to do, see Dana Nelkin, "Two Standpoints and the Belief in Freedom," *Journal of Philosophy* **97** (2000): 564-76.

our choice situation.[12]

Of course, there are requirements of consistency that agent ought to strive to satisfy. But, similarly, requirements of consistency can give me reasons to accept a proposition (or a norm), only if rejecting it would be inconsistent with something else I recognize *I have good reason to accept*.[13] Suppose I can't shake off the belief that **p**, but I admit I have no reason to believe that **p**. Suppose I realize that **p** and **q** are inconsistent. Do I now have a *reason* to believe **not q**? It might be that given my unshakeable belief that **p** makes it psychologically difficult, or even impossible, not to believe **not q**, but it certainly doesn't give me a *reason* to believe **not q**. But if the requirements of consistency fail to generate reasons in this manner in the case of belief, they certainly will do no better in case in which all we need is to act as if we believe that **p**. An example that also seems to involve a case of having to act as if one believes that **p** might help making this point. It seems plausible to say that when one plays a competitive game, one must act as if one believed one could win the game, given that, in many views of games, one can only be playing a competitive game if one is trying to win. In this case, I must act as if I believed that I could win even if, in fact, I believe that I have no hopes of winning. So suppose I am facing Roger Federer in a game of tennis, and although I know I will be crushed I proceed as if I believed I could win the match. Now it is also plausible to suppose that it is a requirement of rationality that if I believe I can win a tennis match against Federer, I should accept certain bets; for instance, a bet that pays me a million dollars if I win and costs me ten dollars if I lose. However, I can certainly face Federer, have many offers for bets with similar odds shouted at me during the match, turn them all down, without thereby being guilty of any kind of irrationality.

We face equally serious problems if we try to understand the requirement to choose as if we were free as a normative requirement; that is, if we read Korsgaard's interpretation as claiming that once one adopts the deliberative standing one is under a *normative demand* to choose as if one were free. If the claim that we must act if we were free is understood as a normative

[12] Allison points out that at most we would need to believe that we are bound by the moral law (see *Kant's Theory of Freedom*: 217); Korsgaard answers a similar charge. But note that the problem I am raising is different; I am arguing that there is no sense in which we even need to *believe* that we are free.

[13] There is a large controversy on how to understand rational requirements, and the conditions under which one can "detach" the consequents of such requirements. See, for instance, John Broome, "Normative Requirements" *Ratio* **12**, 1999: 398-419, and Niko Kolodny, "Why Be Rational", *Mind* **114**, 2005: 509-563. However, the point I am making here is one that I take it all the parties to the controversy would accept.

demand, then on Korsgaard's own view, it must be possible to fail to satisfy this demand. That is, it must be possible that I do *not* choose as if I were free even if I am under the demand to do so. But if it is possible, why shouldn't I? Why is there an obligation to choose in this manner? The argument was supposed to explain the source of an obligation, but it now seems to replace an unexplained obligation by another unexplained obligation. And the explanation of the obligation to act as if we were free cannot be something like "we can't help but see ourselves as free", or "it is an inescapable presupposition of the deliberative standpoint that we are free". This kind of move would send us straight back to the problems of the first interpretation. And just postulating an unexplained obligation as the ground of the moral law would be an instance of what Korsgaard calls "realism" (although I would prefer to call it "dogmatism") in ethics, a view that Korsgaard herself rightly criticizes. But whether or not one accepts Korsgaard's arguments against realism, it does not seem to advance the cause of the moral law to ground it in a further, and possibly less compelling, unexplained obligation.[14]

One might argue that the argument that Korsgaard puts forward here is not very different from the argument that Kant provides, for instance, for the postulates. It seems, for example, that Kant moves, roughly, from the claim that one cannot act from the moral law without believing in the existence of God to a warrant for belief in God. One could thus argue that my argument against Korsgaard's interpretation of Kant's views in *Groundwork III* raises problems for a general argumentative strategy that Kant employs in a variety of contexts. However, if I am right so far, we should also conclude that we should not rush to interpret Kant's arguments for the postulates in the model of Korsgaard's understanding of the necessity to conceive of ourselves as free. Although I cannot get into much detail about Kant's views on the postulates here, a few points may help establish that the arguments I am advancing here are not in tension with Kant's argumentative strategy regarding the postulates. First, at least in the *Critique of Practical Reason*, Kant clearly intends the postulates to be consequences not of something

[14]See, for instance, her *The Sources of Normativity*: in criticizing Prichard, she says "according to Prichard, obligations just exist and nobody needs to prove it" (New York: Cambridge University Press, 1996): 32. Of course, one might think that if one accepts that there is no proof for the moral law in Kant's view, one has put Kant in the same position as realists such as Prichard. Although my argument here does not depend on whether this point is correct, I should point out that it is not. For it is still true on my view that Kant does show that the moral law is an unconditional requirement of practical rationality. So although the moral law is not proven from weaker conceptions of rationality, it is explained in terms of a compelling view of the nature of practical rationality.

such that we cannot help but believe, but of a law that is "apodictically certain" (KpV 47). Kant explicitly says there that the ideas of freedom, God, and immortality "receive objective reality through an *apodictic* practical law" (KpV 135, emphasis mine). So the source of our entitlement to the postulates rests not on something that we must conceive in a certain way by acting, but on something that we are *a priori* conscious of its apodictic certainty. Moreover Kant *does* put severe limits on what can be inferred from the postulates; there is nothing that we learn from the postulates that cannot be inferred directly from the moral law. For instance, despite the fact that postulates give objective reality to concepts that are theoretical concepts,[15] "one can make no theoretical use of them at all." (KpV 135).

It is worth adding that it is not clear that this argument shows that we must regard our choices as free in any interesting sense. Let us assume that the argument does establish that one has to think of one's alternatives as open. But why should we think that this is the same as thinking that one is free? After all, all that it shows is that the alternatives are open with respect to our deliberation; that is, all we need to suppose is that our deliberations, and consequently our choices, are not idle. But to assume that our deliberation and our choices are not idle, all we need is to assume that they are *effective*; we need not assume that they are *free*. That is, I cannot deliberate about whether I should choose to win the New York Marathon, or even better, about whether I should choose to change the laws of nature, because my choice cannot bring about any of these things. Now it is important to note that I am not taking myself to settle any issue about compatibilism (and, I take it, neither is Korsgaard); it might be that various intuitions about moral responsibility and even of some kind of full-blooded agency requires a more or less robust assumption of freedom. All that I am claiming is that the very possibility of deliberation does not depend on any such assumption.

Independently of the issues above, the understanding of freedom presupposed by this interpretation is in serious tension with Kant's view. When we think about freedom in terms of open alternatives, we think in terms that are foreign to Kant's understanding of freedom in his ethical work. No doubt, having the appropriate kind of *external* freedom does involve having alternatives open; in particular, external freedom consists in not having the range of external means available to my power of choice unduly limited by others' choices. But insofar as we're talking about internal freedom, the kind of freedom that consists in having alternatives open seems much

[15] See, KpV 134

like the classic conception of freedom of indifference. Kant considers the conception of freedom that defines freedom as consisting in the capacity of choosing either one alternative or another, the classic notion of freedom of indifference, to be a confused conception:

> But freedom of choice [Willkür] cannot be defined ... as the ability to make a choice [Wahl] for or against the law (*libertas indi erentiae*)... Freedom can never be located in a rational subject's being able to choose in opposition to his (lawgiving) reason, even though experience proves often enough that this happens (though we still cannot comprehend how this is possible). For it is one thing to accept a proposition (on the basis of experience) and another thing to make it the *expository principle* (of the concept of free choice) ... It would be a definition that added to the practical concept the *exercise* of it, as this is taught by experience, a *hybrid de nition (de nito hybrida)* that puts the concept in a false light.[16]

For Kant, freedom is the capacity of self-determination, and it is only through experience that we become aware of the fact that we could choose to act against the moral law. That is, we know through experience that we can fail to exercise the capacity to act from the moral law properly, but the possibility of exercising the capacity of self-determination poorly in this manner is not part of the nature of a self-determining being. But if this is Kant's view, he cannot hold that there is any kind of conceptual connection between choosing and conceiving that one's alternatives are open in the sense that there is more than one alternative that I could end up willing. In fact, for perfectly rational agents, alternatives are *not* and are not conceived to be, open in this manner; perfectly rational beings know that they will always act in accordance with the moral law. It is not part of Kant's concept of freedom that my will could turn in more than one direction, but

[16]MS, 226-227. Korsgaard herself cites this passage, but she seems to think that this understanding of freedom is ruled out by a conception of freedom, rather than by the concept of freedom. I am not sure I know how to import Rawls's distinction between concept and conception into Kant's understanding of freedom (and, in particular, I don't see how Kant's distinction between a negative and positive concept of freedom should be accounted for in terms of the Rawlsian distinction as Korsgaard suggests), but given Kant seems to be implying in this passage that we learn of the possibility of making a choice against the law only through experience, and that he claims in this passage that the possibility of acting against the moral law pertains to our understanding of the *exercise* of freedom, not of freedom itself, it seems that the *concept* of freedom has to be explicated in terms of the ability to follow the moral law.

only that my will, that is practical reason, is effective on its own. A free will for Kant is one whose object (its end) is not given from the outside but is fully determined by its spontaneous activity. Nothing in this notion requires that there is more than one end that reason could set on its own.

Before we move on, I would like to look at an attempt to find an argument in *Groundwork III* for the fact that we are free that does not rely on this kind of understandig of what it is to act under the idea of freedom.[17] Allen Wood tries to understand what is to act under the idea of freedom exactly by trying to understand freedom as a rational relation.[18] To say that an agent acts freely on this view is to say that her actions are explained by norms of reason. Wood goes on to argue that in all our rational judgments we must regard ourselves as following the laws of reason, and thus free. Of course within a certain understanding of "rational judgment" this is somewhat trivial; if we understand a "rational judgment" to be defined in terms of a judgment in which we regard ourselves as following the laws of reason, then the claim above is true, but it will not carry us very far. More fruitfully, we can say that whenever engaged in some kind of enquiry, we must see ourselves following the norms of reason (and, of course, we must take care to understand the notion of "enquiry" here in such a way as not to be one that would again make this claim tautological). We might think (although I don't think this is indisputable) that while engaged in theoretical enquiry understood this way, we are in some way bound by the norms of theoretical rationality. And it might even be plausible, though surely not uncontroversial, to say that insofar as we are capable of engaging in theoretical enquiry we are bound by these laws in the formation of our attitudes that represent the world. And here it seems that we can conclude that insofar as we are engaged in *acting*, or at least in trying to act rationally or something like that, we are bound by the norms governing rational action, and *a fortiori*, by the law that governs the rational choice of end; namely, the moral law.

But even if we accept all this, it will not lead us to the desired conclusion. The moral law is a rational norm for beings that are capable of being motivated by reason alone. Nothing in this argument shows that we must conceive of ourselves in this way. What would it be to show that we are,

[17] I deal more generally with attempts to find an argument from the fact of freedom to the moral law later in *Groundwork III* in section 4.

[18] In fairness to Wood, he describes his interpretation of Kant's argument that I present here as an "unashamed reconstruction and deliberate simplification" (*Kant's Ethical Thought*: 171). But since I find his reading to be a subtle and plausible (though I hope to show ultimately incorrect) reading of the text, I'll proceed to disregard this warning.

or at least must conceive of ourselves to be, beings of this kind? As I understand Kant, what marks beings of this kind is that they take an interest in these rational commands.[19] Kant, of course, does think that we take an interest in these rational commands, but this interest manifests itself exactly in our commitment to the moral law. It's the fact that our interest in these rational commands can only be demonstrated via our commitment to the moral law that leads Kant to the suspicion that the opening arguments of *Groundwork III* can only lead us to a circle. Here is what Kant says:

> Why, then, ought I subject myself to this principle...? I am willing to admit that no interest *impels* me to do so, ... but I must necessarily *take* an interest in it... It seems, then, that in the idea of freedom *we have actually only presupposed the moral law* (G 449, last italics added).

In other words, in asserting that we take an interest in these rational commands we presupposed our awareness of the moral law, and thus any argument that tries to make use of this fact has also thereby presupposed the moral law and can provide no independent access to it.[20]

3 Under the Idea of Freedom

It is tempting to read "idea of freedom" here as "belief that we are free" or "conceiving ourselves to be free". This is not groundless as far as Kant's use of the word "idea" is concerned; an idea for Kant is a concept of reason, and thus it seems that Kant could as well have said "under the concept of freedom", except that "idea" is the more specific word in the case of freedom. So if we follow this seemingly straightforward understanding of Kant's use of "idea", we would have to read the phrase "act under the Idea of freedom" as an ellipsis that should be spelled out as "the assumption that such and such is the case" or "just as if we were free". Although this might seem straightforward enough, the more we look into Kant's use of the word "idea", the less this seems like a plausible interpretation of the expression. Kant does use "idea" also to mean something more like what we commonly refer by the word "ideal". In the *Critique of Pure Reason*, Kant talks about

[19] Cp. "From the concept of an incentive arises that of an *interest*, which can never be attributed to a being unless it has reason and which signifies an *incentive* of the will insofar as it is *represented by reason*. Since in a morally good will the law itself must be the incentive, the *moral interest* is a pure sense-free interest of practical reason alone." (KpV, 79)

[20] I come back to some of these issues in section 4.

ideas being some kind of archetype. Kant relates his use of the word to Plato's in claiming that he is using the expression roughly as the Greek philosopher had done, but pointing out that sometimes "we understand him [a philosopher] even better than he understood himself"(KrV, B 370). Kant explicitly compares ideas and ideals in one of his Lectures on Ethics, and the difference there between the two lies solely on how a certain standard is used:

> We require a standard for measuring degree. The standard may be either natural or arbitrary, according as the quantity is or is not determined by means of concepts a priori. What then is the determinate standard by means of which we measure quantities which are determined a priori? The standard in such a case is the upper limit, the maximum possible. Where this standard is used as a measure of lesser quantities, it is an idea; when it is used as a pattern, it is an ideal. (VE 208 (202))

Although these *Lectures on Ethics* predate the critical period, the claim that Kant makes there about an idea as some kind of standard echoes what he says in the *Critique of Pure Reason* about the idea of virtue:

> We are all aware that when someone is represented as a model of virtue, we always have the true original in our mind alone ... But it is this that is the idea of virtue, in regards to which all possible objects of experience do serve as examples ... but never as archetypes. That no human being will ever act adequately to what the pure idea of virtue contains does not prove in the least that there is something chimerical in this thought... and so this idea necessarily lies at the ground of every approach to moral perfection. (KrV 372)

If we take this understanding of "idea" as our guide, we can say that "acting under the Idea of Freedom" means to act under a certain kind of ideal of a certain kind of perfection.[21] The perfection in this case is the unlimited

[21] In a recent book, Wood also points out that "idea of freedom" should be understood in terms of something like a standard or ideal. Wood, however, glosses it as "any norm that is self-given by reason" (*Kantian Ethics* (New York: Cambridge University Press, 2007): 130), which in his account, would include also the norms of theoretical rationality. Wood uses this gloss to read the argument from *Groundwork III* as moving from the claim that even in theoretical reason we act under the idea of freedom to the claim that we are bound by the moral law. But I do not see how this would work. Wood is trying to move from the fact that we recognize the validity of the norms of theoretical reason to the validity of

use of reason, not "unlimited" in the sense of being infinitely powerful, but in the sense of having no external limitations. If freedom is understood as self-determination, and the relevant limitation here is a susceptibility to sensible incentives, acting under the Idea of freedom is having as an ideal pure self-determination, the ideal of being determined by practical reason alone without the motivating influence of sensible impulses. This understanding not only accounts for what Kant thinks it follows from the fact that rational agents act under the Idea of Freedom, but also meshes well with the explanation that Kant himself gives of why all rational agents must act under the Idea of Freedom.

Let us start with the latter. When Kant tries to explain why rational agents must act under the idea of Freedom he says that:

> for in such a being we think of a reason that is practical, that is, has causality with respect to its objects. Now, one cannot possibly think of a reason that would consciously receive direction from any other quarter with respect to its judgments, since the subject would then attribute the determination of his judgement not to his reason, but to an impulse. (G 448)

Here it might be worth examining what Kant means by a reason that is genuinely practical. Reason, insofar as it is practical, is capable of determining a subject to action. As such, it need not receive any kind of motivational aid from our sensible impulses. But if reason is guiding us correctly it must do it under an ideal of self-determination; it must not relinquish its control to sensible impulses, otherwise it would not be reason itself determining the action, but impulse. This does not imply that a rational being could not deliberate in such a way that she relinquishes control of her actions to sensible impulses, or even that relinquishing control in this way might be the only way in which a rational being could deliberate. It only implies that this would not be a being for whom reason was practical *with regard to its ends*. From all we've said so far there could be Humean beings who are capable of employing reason in the service of passion in exactly the way described in the footnote of the *Religion* quoted above. These beings would not have an independently practical reason; that is, reason would not determine them to act on its own, but rather they would be beings for whom Hume's description would be apt: beings for who reason is a slave of the passions, or such

the moral law. But even if we can connect our recognition of the validity of the norms of theoretical reason to the idea of freedom, if the idea of freedom is understood simply as "any norm that is self-given by reason", why would this imply a commitment also to the moral law, which is a *different* norm of reason? I return to similar issues in section 4.

that reason is merely at the service of the sensible determination of their actions.

However, beings whose reason is capable of determining them to action on its own must be guided by an ideal of self-determination, an ideal of motivation effected solely by rational incentives. And this is exactly what Kant wants to say: every rational being with a will (i.e. practical reason) must act under the idea of freedom. Now what is supposed to follow from the claim that we must act under the idea of freedom? According to Kant such a being is free in a practical respect. "Free in a practical respect" can sound deflationary, as being something qualified or less committal than free *simpliciter*, so ideally we would try to understand precisely what is implied by the restriction "in a practical respect". But all that matters to us is that Kant seems to equate being "free in a practical respect" with being someone for whom "all laws that are inseparably bound with freedom hold for him just as if his will had been validly pronounced free also in itself and in theoretical philosophy". And this seems correct; if a rational being is committed to the ideal of unlimited self-determination, then it is bound by all laws that a fully self-determining being would follow. The laws of freedom are simply the specification of the ideal of self-determination; they simply tell us how an unlimited self-determining being would act. The point is not that rational agents must act *as if* they are free, and thus are bound by the laws of freedom. Agents whose reason is practical on its own *are* free and thus genuinely bound by the moral law. This argument, however, is completely neutral on the nature of our epistemic access to the fact that we are this kind of self-determining being; it is thus silent on the question of whether our awareness of the moral law or awareness of freedom is primary. If my reading is correct, a being to whom reason is not practical on its own is a being for whom we cannot strictly speaking ascribe a will (*Wille*); such a being might have choice *Willkür*, and a faculty of desire, but not a will. If the Humean beings above are possible, they would be beings who, to use Kant's later descriptions, would be capable of bringing about the object of their faculty of desire through choice, but not beings for whom reason could determine on its own the actual object of the faculty of desire.[22] Although the distinction between *Wille* and *Willkür* is not clearly articulated until later work, this understanding of *Wille* is certainly in line with the definition of the *Metaphysics of Morals*:

> The faculty of desire whose inner determining ground, hence even what pleases it, lies within the subject's reason is called the *will*

[22] See MS 213.

(Wille).

One may insist that we know that we are rational beings with a will, not just a faculty of desire. Doubtless, this is true. But the question is whether our awareness of having a will is independent of our awareness of the moral law. And nothing that Kant says in the opening passages of *Groundwork III* establishes that we have any other kind of access to our awareness that reason can determine us to act on its own.

One might argue that now we face a problem to the one faced by other interpretations of *Groundwork III*. Having rid Kant from a dubious philosophical argument, we might have left him with no argument whatsoever. On the reading of *Groundwork III* I am proposing, Kant would be just obscurely putting forward the claim that self-determining beings are indeed self-determining. Now I will argue in a moment that the claim that we must act under the idea of freedom is not supposed to establish any major conclusion of the argument (not even preliminarily), but to set up the proper issue to be resolved: how a finitely rational being could be bound by the laws that apply to an infinitely rational being. It is no surprise then that this step does not provide us with major substantive conclusions. However, Kant does derive an important consequence from this line of reasoning; namely, the claim that we are excused from finding a theoretical proof of freedom. It is Kant's conception of morality as the realization of autonomy that dispenses with the need of establishing freedom from a theoretical standpoint. Heteronomous systems of morality cannot appeal to the Reciprocity Thesis and thus face a double task. First, they need to establish what is the content of our obligation. So, for instance, a certain version of rationalism will advance a conception of perfection as the end of morality. But even if we agree that reason can determine what this conception of perfection is, and that we have a practical interest in the pursuit of perfection, it would not thereby establish that in having perfection as our aim (or failing to have it) we are acting freely and thus it would not establish that our moral (or immoral) actions are attributable to us. Since, *ex hypothesis* our interest in perfection is not necessarily connected to the will's power of self-determination, the only way we can establish our freedom in this picture is by means of a theoretical proof. Or, in other words, heteronomous systems of morality do not have a positive conception of freedom; they do not have a conception of the will being capable of "being a law to itself" (G 447), or of practical reason having its own laws just in virtue of being reason's capacity to determine itself.

Thus, it makes no sense in these systems to claim that a rational being acts under the idea of freedom, since there's no such thing as an ideal, or a perfection, that is implied by the very notion of a self-determining being. And thus, unlike Kant, advocates of such systems cannot move from this idea to the realization that theoretical grounds for freedom are needless. It is only because we can connect rational agency with the idea of freedom that we need no theoretical proof of freedom. Given Kant's conception of rational action as autonomous action, acting from the moral law and acting freely are not two different things. In acting from the moral law, I also act in accordance with the ideal of self-determination or freedom, or, as Kant will later say, in being most virtuous I am also most free. Since the practical laws are the laws of freedom, insofar as we are rational agents, insofar as act from the moral law, we *are* free. In other words, insofar as we can establish the moral law is valid and ought to guide our action, we need not provide a further theoretical proof of our freedom.

Of course, if I am right, the claim that we must act under the idea of freedom presupposes, rather than establishes, that human beings are these kinds of rational agents; if I am correct that Kant did not change his mind on the cognitive primacy of the moral law, Kant must recognize even in the *Groundwork* that this claim is established through our awareness of the moral law, rather than being an independent way to gain access to the moral law. However, whatever the "point of entry" is in the reciprocity thesis, we do know that a proof of theoretical freedom is unnecessary. If the moral law is the law of rational agency, and if rational agency must be self-determining, then awareness of being a rational agent of this kind, in whatever form, suffices to make the idea of freedom something that I must act under, not simply because such an awareness forces me to conceive myself as free, but because such awareness implies, or more precisely, *is* the awareness of my freedom.

4 The Hidden Circle

So far, all I have done is to show that the claim that we act under the idea of freedom on its own does not give an entry way into our commitment to the moral law. Not all Kant interpreters who claim that Kant is trying to infer our commitment to the moral law from the fact that we're free in the *Groundwork* think that the job is mostly done by these "preparatory

arguments".[23] Some interpreters see most of the weight of the derivation as coming later. After all Kant seems to complain just after the passages that we've been focussing on that the argument has been moving in circles. As Kant says:

> It must be freely admitted that a kind of circle comes to light here from which, as it seems, there is no way to escape. We take ourselves as free in the order of efficient causes in order to think of ourselves under moral laws in the order of ends; and we afterwards think of ourselves as subject to these laws because we have ascribed to ourselves freedom of the will. (G 450)

Kant goes on to say that because freedom and the moral law are reciprocal concepts "one cannot be used to explain the other or to furnish a ground for it". However, Kant goes on to say later that the suspicion of a circle can be set aside, after he mentions how the ideas of reason, even in their theoretical reason, lead us to see ourselves as intelligible beings. Here is what he says:

> The suspicion that we raised now above is now removed, the suspicion that a hidden circle was contained in our inference from freedom to autonomy and from the latter to the moral law, namely ... that we were ... unable to furnish any ground at all for the moral law but could put it forward only as a *petitio principii* [*Erbittung eines Prinzips*] disposed souls would gladly grant us, but never as a demonstrable proposition. [G 453]

It is hard not to see these sections as completing an argument that moves us from freedom to the moral law exactly as I have been denying; an argument that breaks the circle of the Reciprocity Thesis by establishing our freedom. And given that between the statement of the suspicion of the circle and the above conclusion that the suspicion has finally been removed, the Groundwork provides a discussion of the role of reason in theoretical inquiry, it is obviously tempting to assume now that Kant is finding in the investigation of theoretical reasoning grounds to establish that we are free without relying upon our commitment to the moral law. Now I take the crucial passage for such a reading to be the following:

> Now a human being really finds in himself a capacity by which he distinguishes himself from all other things, even from himself

[23] The label "preparatory argument" is from Allison. See *Kant's Theory of Freedom*: 214.

> insofar as he is affected by objects, and that is *reason*. This, as pure self-activity is raised even above the *understanding* ... reason ... shows in what we call "ideas" a spontaneity so pure that it thereby goes far beyond anything that sensibility can ever afford it ... Because of this a rational being must regard himself *as intelligence* ... as belonging not to the world of sense but to the intelligible world (G 452).

Given that the ideas of reason play at least a regulative role in theoretical reason, it seems natural to think that, somehow, reflection on our nature as theoretical knowers suffice to commit us to the moral law. Now in discussing Wood I have already tried to expose the difficulties of at least one attempt to move from the nature of our engagement in theoretical inquiry to the conclusion that we are free. But given this kind of textual evidence, one might wonder whether other attempts wouldn't fare better. Although I can't try to foreclose every possible interpretation of how this move could be made, it is worth pointing out problems that would face any such interpretation. Let us ask exactly how we are supposed to understand the role of the ideas of reason in the argument. It would be obviously implausible to ascribe to Kant the view that merely by grasping an idea of reason, in merely representing, for instance, God or a free being, we are already behaving in accordance with the moral law. Not much less implausible would be the claim that merely in representing God we are *immediately* aware of our obligation to obey the moral law. In other words, awareness of ideas of reason in general cannot be simply *identified* with awareness of the moral law. A plausible interpretation would claim that the "pure spontaneity", or "pure self-activity", that Kant is talking about here is *not* the exercise of a capacity to obey the moral law. Consequently, such an interpretation must postulate awareness of some kind of pure spontaneity that is not simply identical to awareness of the moral law, but at the same time forces us to regard ourselves as members of the intelligible world (and thus as free). But this is a puzzling step. For if there is such a thing as pure self-activity that is not acting from the moral law, but is instead, say, following some other kind of rational principle, how could it force us to regard ourselves as member of the intelligible world and, more specifically, as free agents?[24] This argument goes through only if being subject to such rational principles would require us to see our actions as determined by non-empirical laws, and, in particular, by the laws of freedom. But the moral law *is* the law of freedom. And since

[24]For similar concerns about similar readings, see Allison, *Kant's Theory of Freedom*: 217-8.

morality and freedom are reciprocal concepts, there could be no laws of freedom that are not identical to (or at least a consequence of) the moral law. This is made relatively clear by Kant in *Groundwork III* itself:

> With the idea of freedom the concept of *autonomy* is now inseparably combined, and with the concept of autonomy the universal principle of morality, which in idea is the ground of *all* actions of *rational beings*, just as the law of nature is the ground of all appearances. (G 452-3; middle italics added)

But suppose the argument did establish that there was some kind of theoretical activity that entitled us to see ourselves as members of the intelligible world. It does not follow from the fact that a being is a member of the intelligible world that this being is a free agent, and thus it does not follow from the fact that a being is a member of the intelligible world that it is bound by the moral. To show that this is the case one would have at least to show that such a being has a pure will; that is, one would have to show that, specifically, its *faculty of desire* had a corresponding intelligible aspect (what Kant calls a "higher faculty of desire"), or, as Kant puts it later in *Groundwork III* that it has "a faculty distinct from a mere faculty of desire".[25] But this does not obviously follow from the fact that we have the capacity of have ideas of reason even if it is true that such ideas of reason places us in the intelligible world. Of course, there could be an argument that takes exactly this missing step, and, although I have no idea how one could make this move, I don't want to rule out in advance the possibility of such an argument. However, there's no indication of such an argument in these passages; Kant seems to move immediately from our belonging to the intelligible world to the claim that we must act under the idea of freedom and the claim that the moral law is "in idea" the law of all our actions.

One might also put forward a weaker claim; one might want to say that if there is a form of theoretical activity that places us in the intelligible world, then we have taken a step in the right direction even if we have not strictly proven the validity of the moral law. After all, we would have shown that we are already committed to the possibility of our intelligible selves (or an intelligible aspect of our selves) determining our cognition (in this case the ideas of reason). We would then have established exactly what needs to be shown to be possible in the case of the moral law: that an intelligible aspect of our selves (our free will) determines a cognition (our awareness of our obligations, and our actions that are determined by this awareness).

[25] G 459, see also G 461.

This would fall short of establishing the reality of the free will, but at least would show that we have no reason to be skeptical of the validity of the moral law on account of its implications about our nature (that is, that we have an intelligible existence), or of our cognition (that it is determined by our intelligible aspect of ourselves). Although I do not think that this is the correct reading of the passage given the problems with the idea of purely theoretical activity placing ourselves in the intelligible world, this reading of the passage would be congenial to my own interpretation; in fact, as it will become clear later, this reading of the argument would result in a structurally similar interpretation of Kant's deduction of the moral law and of his solution to the "hidden" circle.

Before we move on, I would like to point another important problem for this reading. When we look more closely at the text of the *Groundwork*, it is not so easy to make sense of the idea that Kant thought that he had "broken" the circle in this manner; that is that he thought that it would be possible to find an epistemic route from the fact of our freedom to the validity of the moral law. In particular, if we read the last quote as claiming that we find a way to break the circle, we'd think that the conclusion of the argument that led to this point was something like "we are free", or at least the more problematic "we must assume that we are free". But the recapitulation of the conclusion just after the quote above does not support this interpretation. Kant says:

> For we now see that when we think of ourselves as free we transfer ourselves into the world of understanding as members of it... but if we think of ourselves as put under obligation we regard ourselves as belonging to the world of sense and yet at the same time to the world of understanding. (G 453)

This passage neither says, nor even seem compatible, with the claim that we break the circle by finding a more satisfactory proof of the fact that we are free. In fact, this recapitulation suggests that we are *contrasting* thinking of ourselves as free with something else; namely putting ourselves under obligation. This would be a surprising way, to say the least, to describe a conclusion to the effect that we have found an independent proof of our freedom. This might suggest that the suspicion of moving in a circle is removed not by finding that we can prove one of the sides of the biconditional without assuming the other, but by finding a different way to establish that the moral law is a demonstrable proposition. It might also suggest that Kant might be aiming for something different than what one might expect from the phrase "to put the moral law forward as a demonstrable proposition".

After all, what is put forward as supporting the claim that we escaped the circle does not seem like anything that could function as the materials for a deductively valid proof of the moral law from more secure foundations. More particularly, the passage seems to suggest that what we achieved in the preceding paragraphs was a better understanding of the relation between acting under the idea of freedom and being aware of ourselves as being under obligations.

But is there a plausible reading of these passages that can avoid these shortcomings? We can start by seeing if it is plausible to deny that Kant is saying here that in merely *having* ideas human beings already demonstrate the kind of a pure spontaneity that is unmatched by anything that the understanding can provide. Rather Kant's claim would be that the *content* of these ideas represents the *possibility* of a kind of activity that is pure self-activity. This kind of pure self-activity is what we represent when we represent the capacity of reason to fully determine itself independently of anything that is given to sensibility. So insofar as we think ourselves as rational beings, as beings whose causality can be determined by reason alone, we must think of ourselves as members of the intelligible world. It is important to note that this obviously does not settle the issue of how we know that we are beings of this kind, that we are beings whose causality can be determined by reason alone. The passage does not settle whether we know that we are beings of this kind by being first aware of our freedom or by being first aware of the moral law. Consequently, this passage does not purport to settle the question of the *ratio cognescendi* of the moral law and freedom. So if this interpretation were correct, these passages of the *Groundwork* are fully compatible with the claim of the *Critique of Practical Reason* that the moral law is the *ratio cognescendi* of freedom.

But how could these passages help us solve the problem of the circle under this interpretation? Many interpreters have noticed that the possibility of being members of both these "worlds",[26] or occupying two standpoints, is crucial to how Kant thinks that he can remove the suspicion of moving in circles.[27] However it is also important to point out that the possibility of seeing ourselves as belonging to both an intelligible world and a sensible world is presented in this passage as explaining how we put ourselves *under*

[26]Nothing I say here depends on any particular understanding of the ontology of noumena and phenomena.

[27]Including, of course, Korsgaard herself. Korsgaard, however, has a different understanding of what the circle is, and she wants to claim that these considerations are supposed to explain why there is an "incentive for us to identify with the free and rational side of our nature" ("Morality as Freedom": 165).

obligation. In order to think of ourselves as free, we need to occupy only one world, albeit the most problematic of the two worlds.

My suggestion is to follow the lead that the concept of obligation is the one that Kant finds particularly worrisome in this context: to put forward the moral law as a demonstrable proposition is to show that the concept of an obligation is a coherent one, and more importantly, that knowledge of obligation is possible given the nature of human cognition. After all this is how we cognize the moral law, not as an infallible law of human behaviour, but as the ground of *obligation*. We know that the moral law is the law of an unlimited free will. Such an agent *will* necessarily act as the moral law demands. Insofar as we take an interest in morality we think of ourselves as free, and insofar as we think of ourselves as this kind of agent, the moral laws apply to us; we need no further theoretical proof of freedom. But as finitely rational beings the moral law is for us an imperative; a claim that we *ought* to act as morality requires, even though we know that we do not always so act. The concept of freedom does not contain within itself the concept of an obligation, so even though we need no further proof that a perfectly rational being will act in accordance with the moral law, we still have not explained the possibility of the moral law being valid for a being that is *imperfectly* rational.

The initial circle started because we took ourselves to be free *in the order of e cient causes*, and the Reciprocity Thesis could not explain our entitlement to the idea of a free being operating in the order of efficient causes; it could not explain the possibility of a will that *ought* to act according to the laws of freedom as opposed to a will that *necessarily acts* according to the laws of freedom. The only thing that can explain this possibility, in Kant's mind, is the metaphysics of transcendental idealism; that is, the possibility that we can think of ourselves as governed by the laws of freedom even at the same time as we consider ourselves to be governed by the laws of nature. The argument of *Groundwork III* allows us to put the moral law forward as a demonstrable proposition in the sense that we can show that theoretical reason does not rule out the possibility of an imperfectly rational will, and, more particularly, of a being who at the same time has an empirical existence and for whom the moral law is a determining ground of the will.

An imperfectly rational agent, understood in this manner, must have the exact kind of synthetic a priori knowledge that we conceive ourselves as having in recognizing that we are bound by the moral law. In seeing herself as a member of both the sensible world and of the noumenal world, this kind of imperfectly rational being recognizes that she ought to make her behaviour in the sensible world conform to the laws of the noumenal

world; that is, she must determine herself to act in accordance with rational laws. By understanding the possibility of being a member of both worlds, we understand the possibility of having knowledge of a law that *ought* to govern behaviour, rather than a law that in fact *does* govern behaviour. This might seem like an unambitious conception of what it takes to "put the moral law forward as a demonstrable proposition." But, for Kant, what often stands in need of proof for a priori synthetic proposition is not the content of the proposition, but the possibility of such an *a priori* cognition. And this is exactly what I am suggesting that Kant's deduction of the categorical imperative in *Groundwork III* establishes: how such a practical synthetic a priori principle is possible.[28]

One might now suspect that, under this interpretation, *Groundwork III* accomplishes nothing that has not already been accomplished by the solution to the Third Antinomy. After all, hadn't we learned in the *Critique of Pure Reason* of exactly the possibility that an empirical being could also be a self-determining being? In fact, this is precisely the suggestion I want to make: *Groundwork III* establishes nothing above and beyond what Kant took himself to have established in the solution to the Third antinomy. This would be a disappointing result, however, only if we had reason to suspect that Kant took himself to be doing something different. It might seem obvious that Kant could not be engaged in such a thankless task; after all, why would he not just refer us back to the *Critique of Pure Reason*, rather than produce such a convoluted version of what are essentially the same arguments? However the *Groundwork* is supposed to proceed independently from the *Critique of Pure Reason*, taking as its starting point what is already available to ordinary reason. The arguments for Transcendental Idealism,

[28]This interpretation might seem to run afoul of some well-known interpretation of Kant's notion of a "deduction". According to these views, a deduction must be a demonstration of a certain judgment on the basis of its being a necessary condition for a further claim or fact that we know to be true or to obtain. Although I cannot here discuss in detail what Kant means by "deduction", it is worth pointing out that it accords well with Dieter Henrich's seminal (and to my mind much more plausible) interpretation of Kant's views on deduction. According to Henrich, "the process through which a possession or a usage is accounted by explaining its origin, such that the rightfulness of the possession or usage becomes apparent, defines the deduction." ("Kant's Notion of the Deduction and the Methodological Background of the First *Critique*" in Eckart Förster, *Kant's Transcendental Deductions* (Stanford: Stanford University Press, 1989: 35). For a similar reading of the notion of deduction, see Ian Proops, "Kant's Legal Metaphor and the Nature of a Deduction" *Journal of the History of Philosophy* **41**: 209-229. Although I find Proops's reading of the role of the fact of reason in the Second *Critique* implausible and in tension with some of the things I say here, his understanding of the notion of deduction is fully compatible with the argument I present here.

as well as the statement of the view itself, which are essential to the solution of the third antinomy are not presupposed in the *Groundwork*. In fact the text that just precedes these passages from G 452-3 constitutes an argument for transcendental idealism from the material available to ordinary reason:

> No subtle reflection is required to make the following remark, and one may assume that the commonest understanding can make it ... that all representations which come to us unwillingly [*ohne unsere Willkür*] ... enable us to cognize objects only as they affect us and we remain ignorant of what they may be in themselves so that, as regards representations of this kind, ... we can achieve only cognition of *appearances* never of *things in themselves* (G 452)

This is precisely what we would expect if my reading of the passage were correct. Since accepting transcendental idealism is essential to the solution of the third antinomy, insofar as the argument of *Groundwork III* essentially follows the argument of the Third Antinomy, we must start by establishing the truth of transcendental idealism from the materials available to us in the *Groundwork*.

It is worth looking back at G 452-3 and examine whether this is indeed a plausible reading of these passages. This particularly important because as we read through G 452-3, it seems hard to avoid the impression that Kant is claiming that awareness of the possession of the faculty of reason itself guarantees our intelligible existence and, consequently, the validity of the moral law. However, one gets this impression only when one approaches these passages assuming that Kant is trying to establish the validity of the moral law from some kind of sparser cognitive materials, or so I'll argue. The crucial part of the argument begins as following:

> Now a human being really finds in himself a faculty (*Vermögen*) by which he distinguishes himself from all other things, even from himself insofar as he is affected by objects, and that is *reason*. This, as pure self-activity, is raised even above *understanding* by this: that though the latter is also self-activity and does not, like sense, contain merely representations that arise when we are affected by things ... yet it can produce from its activity no other concepts than those which serve merely *to bring sensible representations under a rule*.

Although this passage clearly implies that in the employment of reason a human being distinguishes himself from all phenomena, it does not say that

we can find this kind of pure self-activity in the theoretical employment of reason. If we do not think that Kant is trying to establish that the use of theoretical reason already somehow commits us to the moral law, there is no reason to think that the "pure self-activity" that Kant refers to is anything other than our awareness of the moral law. Instead, we can say that Kant is here showing that such pure activity confirms the actuality of the possibility raised earlier after Kant argued that even the "commonest understanding" can grasp that cognition of objects through sensible representation is restricted to knowledge of appearances:

> And thus as regards mere perception and receptivity to sensations he must count himself as belonging to the the *world of sense*, but with with regard *to what there may be of pure activity in him* ... he must count himself as belonging to the *intellectual world*. (G 451, second emphasis mine)

So if my interpretation is correct, we first establish that even the commonest understanding "must yield a distinction ... between a *world of sense* and a *world of understanding* (G 451). This is something that is implicit in our ordinary knowledge of objects. But since human beings' knowledge of themselves through inner sense is knowledge of themselves "belonging to the world of sense", in acquiring such empirical knowledge of themselves they must "necessarily assume something else lying in his basis" (G 451). But this raises the possibility that we do have access to our self as belonging to the intellectual world by something other than the representations of inner sense. The next step of the argument is thus to ask whether we find such pure activity in human cognition. But none of this would imply that in searching for evidence of this pure activity, we need to find something that we can gain access to independently of our awareness of the moral law. In other words, given that theoretical reason must distinguish between appearances and things in themselves, we can ascertain that the same distinction can apply to the self and that insofar as we cognize ourselves as pure activity we cognize ourselves as belonging to the "intellectual world" or as a thing in itself. The law of such pure activity is law of self-determination and thus a law that can (and sometimes does) determine ourselves in the world of sense; what we are in ourselves, of course, must also determine how we *appear*. But this makes room precisely for the possibility of a law that is a law for our actions in the world of sense even though it is not a law that *will* necessarily determine our activity as members of the world of sense; in other words, the possibility of being under an *obligation*. If I am right, this is exactly what we are after; that is, we need to show the possibility of a rational being that

is *obligated by*, rather than necessarily following, the moral law. Now we can go back to the passage we quoted earlier:

> But reason, on the contrary, shows in what we call "ideas" a spontaneity so pure that it thereby goes far beyond anything that sensibility can ever afford it.

If we keep this interpretation in mind, we can now spell out the reading I proposed earlier. As I said, Kant is not saying that merely *having* such an idea, or by merely *thinking*, we display this kind of spontaneity. What the ideas of reason show us is a form of *determination* that is a form of pure self-activity, but this self-activity is only actual when one is in fact fully *determined* by such an idea. Of course, the only idea that can fully determine us to act is the idea of freedom. We can conclude at this point that insofar as we act under the idea of freedom we exhibit the kind of spontaneity that does not belong to the world of sense, and insofar as we are aware of this spontaneity in ourselves, we are aware of ourselves as members of the "intellectual world". We break the circle by showing not simply that the Idea of freedom and the moral law are analytically connected, but by showing that in being aware of a capacity for acting under the idea of freedom I am aware of being a member of a world other than the empirical world, and thus being, as it were, legitimately thought to be under the jurisdiction of laws other than the empirical laws; namely, the laws of freedom. This is indeed how Kant represents what he has shown in the following section "How is a Categorical Imperative Possible":

> But because *the world of understanding contains the ground of the world sense and so too of its laws*, and is therefore immediately lawgiving with respect to my will ... it follows that I shall cognize myself as intelligence, though on the other side as being belonging to the world of sense, as nevertheless subject to the law of the world of understanding. (G 453-4)

So if I am right that this is a possible reading of these passages, even the portions of *Groundwork III* that seem most congenial to a claim that Kant was searching for a proof of the validity of the moral law from a thinner notion of rational agency can be understood so as to fit well with the interpretation I propose. I also argued that we have good textual and philosophical basis to avoid the more ambitious interpretation; in particular, I have argued we have good philosophical and textual reasons to conclude that Kant already held the view that the moral law was the *ratio cognoscendi* of freedom

rather than the other way around in the *Groundwork III*, and thus that we have good reason to think that nothing essential changes in his views about the relation between freedom and the moral law between the writing of the *Groundwork* and the Second *Critique*. However there are some differences in the two works that we have not yet addressed that seem to indicate a change in position. So before I can complete the case in favour of my interpretation, it is worth addressing these passages.

5 Does Anything Change from the *Groundwork* to the *Critique*?

Does anything change in Kant's view from the *Groundwork* to the *Critique of Practical Reason*? I have said above that Kant never seems to acknowledge such a change and that all the materials to the conclusion that there could be no independent proof of the reality of freedom seemed readily available to Kant. Yet, there is some evidence that at the very least a change in emphases must have taken place. Two striking contrasts are worth pointing out. First Kant describes the last section of the *Groundwork* as a critique of *pure* practical reason,[29] and yet, Kant quickly asserts in the *Critique* that such an endeavour is unnecessary, and that one should pursue, not a critical examination of *pure practical reason*, but simply of *practical reason*. As Kant puts it:

> If we can now discover grounds for proving that this property [freedom] does in fact belong to the human will, then it will not only be shown that pure reason can be practical, but that it alone ... is unconditionally practical. Consequently, we shall not have to do a critique of *pure practical* reason but only of *practical reason* as such. For, pure reason, once shown to exist, needs no critique. (KpV 15-16)

Secondly, and perhaps more importantly, while the section "On the Deduction of the Principles of Practical Reason" in the *Critique of Practical Reason* comes to the conclusion that there can be no deduction of the moral law, [30] Kant seems to think that he succeeded in delivering such a deduction in *Groundwork III*. In fact, the section ends with Kant happily referring to

[29] See also, G 391.

[30] Kant famously refers to the deduction of the moral law in the Second *Critique* as "a vainly sought deduction" (KpV 47).

what he had done as the "deduction of the supreme principle of morality" (G 463).

Questions about what Kant means by "deduction" and "critique", as well as what would be difference between a critique of a pure faculty as opposed to the critique of the faculty in general, raise large interpretative issues that I cannot settle here. However, I think we already have the materials to explain these changes from the *Groundwork* to the *Critique of Practical Reason*. Let us start with the second issue. As I said above, the *Groundwork* does not presuppose the arguments of the first *Critique*; its starting point is ordinary reason, not Kant's previous speculative accomplishments. As such, it cannot presuppose that there is no conflict between the demands of the categorical imperative and the laws of the natural world. In fact, the only other passage in *Groundwork III* in which Kant refers to what he has accomplished as a "deduction" is at the conclusion of the section "How is a Categorical Imperative Possible?" (G 454). "Deduction" there refers, strictly speaking, to a deduction not of the moral law, but of the categorical imperative; in particular, it is a deduction of the possibility of the moral law binding *nitely* rational beings like us. But this kind of deduction is superfluous if one can simply avail oneself of the results of the *Critique of Pure Reason*. The "vainly sought deduction" of the moral law in the *Critique of Practical Reason* is not a deduction of the possibility of the categorical imperative, but of the *objective reality* of the moral law. The vainly sought deduction is supposed to be a deduction of whether the moral law is in fact capable of determining our will. It would be a proof that it is within the power of my will to follow the moral law; in other words, not only a proof that the moral law can be the law of the will of a finitely rational being, but a proof that *I* am capable of following its commands, a proof that proceeds independently of my awareness of the moral law. But no such proof is possible or necessary. I have no independent epistemic access to my freedom (i.e. to my capacity to follow the moral law), and the apodictic certainty provided by my awareness of the moral law guarantees that I am capable of obeying its commands (since "ought" implies "can").[31] Similar considerations can explain Kant's change in view with regard the need of a critique of *pure* practical reason. If we think that the critical project is concerned at least in part with the boundaries of our faculties, and more particularly, with the possibility that our cognitive faculties might overstep their bounds, we can see that the critical concerns of the *Groundwork* and the *Critique of Practical Reason* will be somewhat different. The *Groundwork* needs to

[31] See KpV 47.

show that the exercise of practical reason does not make assumptions that conflict with our theoretical knowledge of nature. In this context, it would be particularly important to show that *pure* practical reason, that is practical reason when considered independently of anything given from sensibility, does not presuppose that the will can be determined in a way that is incompatible with the determination of our actions by the law of nature. It is, after all, the pure use of practical reason that purports to determine our will by its own laws.

On the other hand, we start the *Critique of Practical Reason* knowing that no such conflict is possible, knowing that the theoretical use of reason leaves open the question of whether there could also be a causality of freedom determining our will. In the *Critique of Practical Reason* we examine the scope and nature of the principles of practical reason already in the knowledge that the practical employment of reason does not conflict with its theoretical employment. However, in theoretical reason, the pure employment of our cognitive faculties raises a further concern regarding the application of the principles of understanding to objects. Since understanding does not produce its own objects, a critique is needed in order to establish whether (or in which cases) the understanding oversteps its boundaries in applying its principles to objects.[32] However, in the case of practical reason there could be no further concern whether practical reason oversteps its boundaries when it is applied to its object; the pure use of practical reason is concerned simply with *self*-determination, with determination of the will itself. The object of the pure use of practical reason is the will itself, so there can be no issue of its legitimate application to objects.[33] In sum, there can be a concern that the pure employment of practical reason oversteps its boundaries only in relation to a possible conflict with theoretical reason, but not in relation to a possibly illegitimate application to objects.

Finally, I should note that the arguments of this paper are silent on whether Kant's discussion of practical freedom in the *Critique of Pure Reason* could provide an argument from an independent conception of freedom to the moral law *circa* 1781. However the relation between practical freedom, transcendental freedom, and the moral law in the first *Critique* is a notoriously difficult issue that deserves separate treatment. Still, for our purposes, it is worth pointing out that there is no explicit argument of this

[32]See KpV 15: "A critique of it [sc. reason] with regard to this [sc. theoretical] use really dealt only with the *pure* cognitive faculty, since this raised the suspicion ... that it might easily lose itself beyond its boundaries."

[33]As Kant says: "For, in that, reason can at least suffice to determine the will and always has objective reality insofar as volition alone is at issue." (KpV 15)

kind in the first *Critique*. Some of the most promising routes to construct such an argument seem incompatible even with views clearly held by Kant in the first *Critique*. For instance Kant seems to claim that practical freedom can be known through our experience of being able to resist immediate sensible incentives (KrV A802/B830) and that there could be no practical freedom without transcendental freedom (KrV A534/B562). These claims seem enough to establish a route for an empirical demonstration of transcendental freedom that does not rely on our awareness of the moral law. However, the claim that our transcendental freedom can be cognized empirically seems to be in stark conflict with the Third Antinomy. So an interpretation that would try to resolve this conflict would likely deprive Kant of such a straightforward argument from the fact of our freedom to the moral law. And an interpretation that left the conflict intact would probably see Kant's views expressed in the *Canon* as some kind of remnant from earlier periods, and thus would not establish that Kant would have an argument of this kind from the views endorsed in the *Groundwork*.[34] At any rate, my aim was to show that there is no great reversal within Kant's major ethical works in the critical period; how far back Kant held these views in his career is a question for another occasion.

One might think that if my view is correct, it has the important disadvantage of depriving some of the interest in Kant's mature ethics. After all, if we can't find in Kant's mature work in ethics an argument for the moral law from thinner premises, we seem to have lost an important avenue in trying to understand the rationality of morality. But, on the other hand, one might take away also a source of skepticism about Kant's ethics. For it is understandably hard to believe that this kind of task can succeed, that one can show that by reasoning about Newtonian Mechanics, we are already somehow committing ourselves to the moral law, or that I cannot adopt a deliberative standpoint when, for instance, choosing items in a restaurant menu, without thereby paving an argumentative path that takes me all the way to an obligation to, say, develop my talents. Impressive and interesting as many of these arguments certainly are, it is hard to avoid approaching them with the sense that someone is trying to sell us the Brooklyn Bridge.

[34]See Allison, *Kant's Theory of Freedom*, ch. 3, and Beck, L. W., *A Commentary on Kant's "Critique of Practical Reason"* (Chicago: Chicago University Press, 1960): 190n, for influential attempts to reconcile the *Canon* and the *Dialectic*. Norman Kemp Smith famously defends the "patchwork theory" more generally, and in particular, argues that the views expressed in these different passages are from different times in the development of Kant's view. See his *A Commentary on Kant's "Critique of Pure Reason"* (New York: Humanities, 1962).

It might be the case that we should be grateful to Kant for leading us away from the attempt to derive the moral law from a thinner notion of freedom or rationality. On the other hand, there is obviously much else of interest and controversy in Kant's view in the neighbourhood of these issues. The most conspicuous example in this context is the Reciprocity Thesis itself. It is far from uncontroversial to hold that freedom and the capacity to act from the moral law are one and the same. Trying to understand the relation between reason, morality, and freedom that is implicit in the Reciprocity Thesis can prove to be an interesting endeavour independently of how we understand our direction of access to these notions. Indeed I find Kant's understanding of the relations among morality, reason, and freedom perhaps the most promising aspect of his moral philosophy. But, fortunately, explaining why this is so lies beyond the scope of this paper. [35]

[35]I would like to thank Matt Boyle, Jennifer Nagel, Arthur Ripstein, David Sussman, Helga Varden, Owen Ware, an anonymous referee for this journal, and an audience at the University of Illinois, Urbana-Champaign for very helpful comments on an earlier version of this paper.