The Deductions of Freedom/Morality-as-Autonomy and the Categorical Imperative in Groundwork III and Their Problems

Fernando Rudy Hiller
Stanford University
ferudy@stanford.edu

Abstract

The first objective of this paper is to present an interpretation of *Groundwork III* which aims to establish two main points: first, that Kant offers there a theoretically-grounded deduction (in a Kantian sense) of freedom/morality-as-autonomy; second, that Kant also offers a *separate* deduction of the categorical imperative. Thus, contrary to what several commentators have claimed, *Groundwork III* contains a theoretically-grounded *double* deduction. The second objective of the paper is to examine and criticize in detail one crucial step in these deductions, namely, Kant's inference from the speculative spontaneity of reason to the noumenal existence of the subject as a free will. I show that Kant himself came to reject this inference in the B edition of the *Critique of Pure Reason*, and argue that this explains Kant's rejection, in the *Critique of Practical Reason*, of the deduction of the moral law he previously offered. Thus, contrary to the "reconciliationist" reading, there is indeed a great reversal in the latter work.

Key Words: *Groundwork*, Kant, deduction, moral law, fact of reason.

Recibido: 19 - 06 - 2015. Aceptado: 18 - 09 - 2015.

Resumen

El primer objetivo de este artículo es presentar una interpretación del tercer capítulo de la *Fundamentación* para establecer dos puntos: primero, que Kant ofrece ahí una deducción (en sentido kantiano) de la libertad/moralidad como autonomía a partir de premisas provenientes exclusivamente de la filosofía teórica. Segundo, que Kant ofrece también una deducción distinta del imperativo categórico. El segundo objetivo del artículo es examinar y criticar en detalle un paso crucial en estas deducciones, a saber, la inferencia de la existencia noumenal del sujeto como voluntad libre a partir de la espontaneidad especulativa de la razón. Muestro que Kant mismo rechaza esta inferencia en la edición B de la *Crítica de la razón pura* y argumento que esto explica el hecho de que Kant, en la *Crítica de la razón práctica*, abandona la deducción de la ley moral que había ofrecido previamente. Así, contra la lectura "reconciliacionista", hay de hecho un gran cambio en la última obra.

Palabras clave: Fundamentación, Kant, deducción, ley moral, hecho de la razón.

One of the most striking puzzles in Kantian moral philosophy is Kant's attempt to provide a deduction of the moral law in the third section of the *Groundwork of the Metaphysics of Morals*. This attempt is puzzling for two reasons. First, a deduction of the moral law—and the deduction of freedom from non-practical premises that, I will argue, Kant attempts as a prerequisite of the former—seems to be in tension with certain key points of Kant's critical philosophy. Second, in the *Critique of Practical Reason* (published in 1788, three years after the *Groundwork*) Kant states that "the objective reality of the moral law cannot be proved by any deduction" (*KprV* 5, 47).[1] Moreover, in that work he completely

[1] I quote Kant's works following the canonical Academy Edition (volume and page number) and the standard A/B pagination for the *Critique of Pure Reason*. The abbreviations I employ are also the standard ones: *KrV* for the first *Critique*; *KprV* for the second *Critique*; and *G* for the *Groundwork*.

inverts the strategy of the *Groundwork*, since the objective reality of the moral law it is now said to be what allows a deduction of freedom. In this paper I want to explore the first reason in order to shed light on the second, that is, I want to understand the deductions of *Groundwork III* and their problems in order to understand Kant's radical change of mind in the second *Critique*.[2]

However, there are commentators—including canonical ones such as H. J. Paton (1965) and Dieter Henrich (1998), and more recent ones such as Julio Esteves (2012) and Sergio Tenenbaum (2012)—who deny that in *Groundwork III* Kant attempted to provide a deduction of the moral law from non-practical premises and that, as a consequence, make the striking claim that there is no "great reversal"[3] in the second *Critique*.[4] These authors agree that the essential elements of the doctrine of the Fact of Reason (introduced in *KprV* 5, 31)—according to which rational beings have unmediated conscience of the bindingness of the moral law, which makes a deduction of the latter either unnecessary or impossible—are already, albeit dimly, present in *Groundwork III*.[5]

---

[2] Henry Allison (1990, 201, 214) also describes his project in this way. However, the explanation he offers of the failure of the deduction in *Groundwork III* is very different from mine. See footnote 37 below.

[3] The term was coined by Karl Ameriks (1982, 226), who does defend the occurrence of a great reversal. My interpretation of the latter (but not of the overall argument in *Groundwork III*) is importantly influenced by Ameriks'.

[4] Jens Timmermann (2010) is an interesting case, because he denies that in *Groundwork III* Kant attempted a deduction of the moral law from non-practical premises (p. 82), but at the same time he does admit that there is a great reversal in the second *Critique* (p. 85). Frederick Rauscher (2009, 205-206) also denies a theoretically-grounded deduction in *Groundwork III*, but he doesn't explicitly discuss whether a great reversal occurred or not, although he seems to suggest that it didn't.

[5] Esteves (2012) doesn't claim that the doctrine of the Fact of Reason was already present in *Groundwork III*. Rather, he claims that in this work Kant performed a (successful) deduction of freedom from *practical*—although non-moral—premises and that in the second *Critique* Kant never recanted this deduction (see esp. p. 158). Thus, since he denies both that there is a theoretically-grounded deduction of freedom in *Groundwork III* and the occurrence of a great reversal in the second *Critique*, I include him as one of my targets.

Rather than arguing piecemeal against the different versions of this "reconciliationist"[6] reading, my strategy for both defending the presence of a theoretically-grounded deduction in *Groundwork III* and a great reversal in the second *Critique* will be to present a detailed interpretation and criticism of the argument in the former work. My interpretation will also reveal, against what is contended by Paton (1965, 247) and Henry Allison (1990, 227), that Kant in fact attempted a *double* deduction: not only a deduction of freedom/morality-as-autonomy (or the moral law), but also a deduction of the categorical imperative.[7]

The plan of the paper is as follows: in the brief first section I present Henrich's (1989) illuminating conception of what a Kantian deduction is. In the second section, divided in several subsections, I reconstruct Kant's argument in *Groundwork III* as a theoretically-grounded double deduction of freedom/morality-as-autonomy and the categorical imperative, and explain why they count as deductions in a Kantian sense. Finally in the third section I present two problems that infect a crucial step in Kant's argument and that, given Kant's own critical principles, strongly push in the direction of recanting the deductions in *Groundwork III*.

1 What is a Kantian deduction?

Let me start by considering how to understand the term "deduction" as Kant uses it. As Henrich (1989) explains, when Kant talks about a deduction this term should *not* be understood in the logical sense of a conclusion derived from premises. Rather—and in analogy with a juridical usage in vogue in the eighteenth century—"deduction" meant for Kant a procedure by which the legitimacy (or objective validity) of a concept or principle is investigated by tracing its origin in the activities of pure reason and pure understanding.

As Henrich shows, a juridical deduction was called for when a person's acquired right over a possession (e.g., a house) or an entitlement

[6] This term was, again, coined by Ameriks (2003).

[7] By contrast, Tenenbaum (2012, 580-581) can be read as claiming that, insofar Kant is attempting a deduction in *Groundwork III*, it is exclusively a deduction of the categorical imperative, not of the moral law. McCarthy (1979) and Timmermann (2010) defend the same view. In 2.6 below I explain the distinction between the moral law and the categorical imperative, and also explain why two deductions are performed in *Groundwork III*.

(e.g., an academic title) was challenged. In order to respond to the challenge, an investigation had to be carried out to show that the way in which the right was acquired—its origin—was lawful. In analogy with this procedure, a Kantian deduction has the purpose of showing, against skeptical doubts, that a concept or principle is legitimate. The paradigmatic case is, of course, the transcendental deduction of the categories,[8] by means of which Kant begins to allay Humean skeptical doubts about the legitimacy of concepts such as substance and cause. This counts as a deduction because Hume's challenge is answered by appealing to origins; in this case, the origins of these concepts are sought for in the activities of pure understanding. In effect, the task of the deduction is to show that the categories "relate to objects *a priori*" (*KrV* A 85/B 117) by way of explaining how the possibility of an object of experience depends on the synthesis of the manifold of sensation performed by the understanding through the categories themselves.

As I will argue in what follows, Kant's argument in *Groundwork III* counts as a deduction precisely in the sense explained above: Kant intends to legitimate the *practical* use of the concepts of freedom, moral law, and the categorical imperative by tracing their origin in the faculty of reason and, more specifically, in certain *cognitive* functions of the latter—a strategy that will cause Kant a great deal of trouble, as I will explain in section 3.

2 The deductions of freedom/morality-as-autonomy and the categorical imperative in *Groundwork III*

There should be little doubt that Kant attempts a deduction (or, as I will show, deductions) in *Groundwork III,* given that he himself calls the argument presented there a deduction on three separate occasions (*G* 4, 447, 454, 463).[9] However, as we will see, it is not altogether clear what exactly the object of the deduction is—whether the moral law,

[8] In the first section of the transcendental deduction of the categories Kant begins precisely by noting the juridical origin of the term "deduction" (*KrV* A 84/ B 116).

[9] In the first of these passages Kant talks about a deduction of freedom; in the second, a deduction of the categorical imperative; in the third, a deduction of the supreme principle of morality. In this section I sort out the complexities about how to make compatible these different claims.

freedom, the categorical imperative, or all of them. Kant also emphasizes that there is an important dissimilarity in the strategy of the first two sections of the *Groundwork* with respect to the third: the former are "merely analytic," because their objective is "unraveling the concept of morality generally in vogue" so as to show that "an autonomy of the will unavoidably attaches to it, or rather lies at its foundation" (*G* 4, 445). By contrast, the third section is synthetic, since the goal there is to *prove* that "morality is no phantasm—which follows if the categorical imperative and with it the autonomy of the will is true and absolutely necessary as an *a priori* principle" (idem). In order to prove this, Kant notes, a critique of pure practical reason is required, for only such a critique can show the objective validity of the central practical concepts.

2.1 The moral law as a synthetic **a priori** *proposition*

In section two of the *Groundwork* Kant claims that the hypothetical imperative, "as far as willing is concerned," is analytic (*G* 4, 417); by contrast, the categorical imperative is a synthetic *a priori* practical proposition (*G* 4, 420). In order to understand Kant's project of a deduction of the moral law,[10] it is essential to understand first in what sense the moral law is expressed as a synthetic *a priori* proposition, since the goal of the deduction is precisely to explain its possibility.[11] Kant's own explanation of the syntheticity of the moral law is not altogether consistent throughout the *Groundwork*,[12] so the reader must do some interpretative work. In what follows I present my interpretation.

To understand the sense in which the moral law is a synthetic *a priori* proposition we have to focus not on any of the basic three formulas of the categorical imperative (universal law, humanity, and autonomy) in particular, but on what they have in common. What they have in common

[10] For now I use the terms "moral law" and "categorical imperative" interchangeably (Kant himself is not very careful in distinguishing them). In 2.6 below I explain why this equivalence is not correct.

[11] By contrast, most commentators don't bother to explain in any detail the syntheticity of the moral law. An important exception is McCarthy (1979).

[12] Kant touches the subject of the syntheticity of the moral law in *G* 4, 420; 421 n.; 440; 447; and 454. Since in each of these occasions (especially in the last one) he gives a slightly (though importantly) different explanation of why the moral law is a synthetic *a priori* proposition, the reader must decide which is Kant's considered position.

is the idea that a will, insofar as it is fully rational and regardless of any particular ends it might have, acts in such a way that the maxim of its action has a specific characteristic. Since this characteristic makes the maxim a moral one, we can summarize the synthetic *a priori* proposition that constitutes the moral law as follows: *a rational will necessarily acts according to moral maxims*. Kant insists in several places that this is indeed a synthetic *a priori* proposition because, from the concept of a rational will, is impossible by mere analysis to arrive at the concept of a will that necessarily acts on moral maxims (*G* 4, 421 n., 440). In other words, the concept of a morally good will is not *contained* in the concept of a rational will—which does not mean, of course, that both concepts are not necessarily connected; they are, but not analytically.

2.2 The deductions, step one: linking rationality and autonomy

Just as the *Critique of Pure Reason* has as its main objective to explain how synthetic *a priori* theoretical judgments are possible, the critique of pure practical reason that Kant sketches in *Groundwork III* has as its goal to explain how synthetic *a priori* practical propositions—the moral law and the categorical imperative—are possible.[13] One must take note of an important similarity and an important dissimilarity in the meaning of the question about possibility between the two works. The similarity is that when Kant raises the "How is it possible?" question he seeks "to discover and to examine the real origin of [a] claim and with that the source of its legitimacy" (Henrich 1989, 35). Although Henrich is talking here about claims of knowledge in the context of the first *Critique*, it is clear that the question of the possibility of the moral law is understood by Kant in the *Groundwork* as a question concerning its legitimacy—in this case, its legitimacy as a demand on every rational being. Moreover, in order to answer this question we must seek the origin of such a law in reason, a task for which a critique of the latter is necessary (*G* 4, 440; 445).

The dissimilarity is that the "How is it possible?" question in the first *Critique* asks for the possibility of a synthetic *a priori* judgment (e.g., "Everything that happens has its cause") or a body of such judgments (such as mathematics and pure natural science) constituting objective *knowledge*. In other words, the answer to the "How is it possible?" question

[13] In 2.6 I defend the claim that these are two distinct practical propositions, and so require different deductions.

comes from showing that the judgment or body of judgments under investigation apply *a priori* to objects and, as a result, constitute (or help constitute) knowledge. By contrast, when Kant asks in the *Groundwork* how the hypothetical and categorical imperatives are possible, what he wants to discover is the ground for "the necessitation of the will that the imperative expresses in its task" (*G* 4, 417). In other words, he wants to explain why the imperatives are *normative* for the will.

Now, since the moral law is a synthetic *a priori* practical proposition, Kant starts by posing the general problem that concerns every synthetic proposition, namely, to explain how subject and predicate are connected in it (*KrV* A 9/B 13; *G* 4, 447). Because synthetic *a priori* propositions are necessary, experience is ruled out as that in which subject and predicate are connected. So the problem of accounting for a synthetic *a priori* proposition (be it theoretical or practical) is the problem of explaining what is the "third thing" (*G* 4, 447) or the "unknown = *X*" (*KrV* A 9/B 13) in which the subject and the predicate of such proposition are "bound together" (*G* 4, 447). And the solution is always to search in the faculties of the subject—as the source of such propositions—for this "third thing" which will explain the possibility of synthetic *a priori* propositions.

As we saw above, the synthetic *a priori* practical proposition that needs to be accounted for is: "a rational will necessarily acts according to moral maxims." Thus, the problem resides in explaining what the third thing that connects "rational will" with "necessarily acts according to moral maxims" is. In other words, we need to explain why a rational will is necessarily (although not analytically) a will that acts morally. But we can be more specific about how to understand "acts morally" here: towards the end of *Groundwork II*, and after presenting and discussing the formula of autonomy and its variant (the formula of the kingdom of ends), Kant sums up by saying that "Autonomy of the will is the characteristic of the will by which it is a law to itself" (*G* 4, 440) and then declares that "the envisaged principle of autonomy is the sole principle of moral science" (idem). Thus, what needs to be explained here is the necessary connection between a rational will and autonomy.[14]

[14] Given Kant's initial definition of the will as a "kind a causality of living beings in so far as they are rational" (*G* 4, 446), it is relatively easy to see why the connection between the concepts "rational will" and "autonomy" is not, as in the case of the hypothetical imperative, analytic: in the latter case, we can say that, since the very concept of a causality involves the production of effects, a

It is patent from the very beginning of *Groundwork III* that Kant is working with the conception of morality as autonomy provided by the third formula of the imperative: in effect, the first subtitle of this section reads "The concept of freedom is the key to the explanation of the autonomy of the will" (*G* 4, 446). So Kant's strategy is clear: if he can show that the rational will is free, he would have explained why the rational will is autonomous and so subject to the moral law. In this subsection Kant sketches very quickly how the transition from freedom to autonomy would go. The starting point is the definition of the will as a "kind a causality of living beings in so far as they are rational" (*G* 4, 446), followed by the postulate that freedom "would be that property of such a causality, as it can be efficient independently of alien causes *determining* it" (idem). Kant makes it clear that the alien causes he has in mind are natural causes, presumably as expressed in our inclinations and desires, since he contrasts freedom of the will with the natural necessity that governs non-rational beings.

As innocuous as this *negative* conception of freedom[15] might seem, Kant thinks it is actually sufficient for showing that a rational will that is free in this negative sense is necessarily (and analytically)[16] governed by the moral law. Kant's argument goes as follows: a will that is independent

---

will that is completely indifferent regarding the production of effects (which is the same as saying that is completely indifferent regarding the means to its ends) is like a causality that cannot cause anything—and this is contradictory. (The negation of analytic propositions results in a contradiction, and this is why we can say that the hypothetical imperative is analytic.) But things are different concerning autonomy. Kant claims that "the concept of causality carries with it that of laws" (idem) linking cause and effect and so, given that the will is "a kind of causality", we can say by analysis that the concept of rational will incorporates that of laws. But what we cannot say by mere analysis is that the concept of rational will (as a causality) incorporates that of *self*-legislated laws, since it is not contradictory to think of a causality that is not autonomous, i.e., that operates according to laws that come from somewhere else (Kant thinks that this is precisely the case with non-rational animals [idem]). Hence, since there is no contradiction here, the link between the rational will and autonomy must be synthetic and, since it is necessary, it must also be given *a priori*.

[15] Negative because it states what freedom is *not*, namely, a causality determined by alien or natural forces.

[16] See footnote 32 for an explanation of why the transition from negative freedom to morality as autonomy is analytic.

of natural causes is *eo ipso* independent of natural laws; but, since it is a causality, the will must have some law or other (*G* 4, 446).[17] But we know that the law that governs the will cannot be a natural law, and that means that the will cannot be heteronomous, since natural necessity (the necessity that accords to natural laws) is a "heteronomy of efficient causes" (idem). What this means is that an efficient cause governed by natural necessity (a rock, say) produces an effect (the breaking of a window) provided that something else, in turn, causes the efficient cause to act (a person throwing the rock). In short, efficient causes in the natural world never determine themselves to action, but always stand as an effect of some other efficient cause, and this is why the necessity that governs their action is heteronomy. But since the will is not governed by natural laws—if it is not heteronomous—that means that it must be governed by its own laws—it must be autonomous. Kant's reasoning here is simple: the laws that govern the will must be either self-given laws or laws imposed on it; they cannot be the latter (since we start from the assumption that the will is negatively free), so they must be the former. But precisely this idea of the will being governed by its own laws, or the will being a law to itself, is the principle of autonomy that in *Groundwork II* Kant presents as the "sole principle of moral science" (*G* 4, 440). Given all this, Kant is able to conclude: "a free will and a will under moral laws are one and the same" (*G* 4, 447).

Hence, starting from the negative conception of freedom as independence from natural laws, Kant arrives at a *positive* conception of freedom, that is, at a substantive conception of what freedom of the will *is*, namely, autonomy. If this (analytic) argument works, Kant has showed us how to (synthetically) link the concepts of a rational will and an autonomous will, namely, *through* the concept of freedom. However, this does not mean that the concept of freedom *itself* is the elusive third thing in which, so to speak, rationality and autonomy meet;[18] rather, Kant claims that the "*positive* concept of freedom *provides* this third thing" (*G* 4, 447, second emphasis added). Not without suspense, Kant adds: "What this third thing is, *to which freedom points us*, and of which

[17] This is stated here just as an assumption, but given the arguments in the first *Critique* aiming to show that the causal connection of events in our experience is always a necessary connection, i.e., a connection according to laws (see for example *KrV* B 234; also A 216/B 263), the assumption is not unmotivated.

[18] Paton (1965, 244) incorrectly makes this assumption.

we a *priori* have an idea, cannot yet at once be indicated here … but still requires some preparation" (idem, emphasis added). Still more intriguing, in the bit omitted from the passage just quoted Kant adds as a parenthetical remark that the mysterious third thing can "make comprehensible the *deduction* of the concept of freedom from pure practical reason, and with it the possibility of a categorical imperative" (idem, emphasis added). So now we have a puzzle: how can freedom disclose the third thing that links rationality and autonomy and, at the same time, be deduced from it? To answer this puzzle we have to recall the specific sense in which Kant uses the term "deduction." This will also shed light on Kant's strategy for deducing the moral law.[19]

2.3 The deductions, step two: the intelligible world and the deduction of freedom

In section 1 we saw that a Kantian deduction is an investigation into the origins of a concept or principle with the purpose of legitimating its use, and that such origins are sought in the subject's cognitive and practical capacities. This sense of deduction is obviously different from the ordinary sense of deduction as the inference of a conclusion from premises according to logical laws. In the latter sense, it would clearly make no sense to say that freedom points to the third thing in which rationality and autonomy meet and that, in turn, this third thing plays a role in deducing freedom. This would be like saying that freedom

[19] The reader may find it perplexing that I skip entirely the subsection of *Groundwork III* entitled "Freedom must be presupposed as a property of the will of all rational beings," where Kant presents the so-called "preparatory argument" in which he argues that "every being that cannot act otherwise than *under the idea of freedom* is actually free, in a practical respect" (*G* 4, 448). My reason for doing so is that, as Allison (1990, 214-218) has argued, the deduction of freedom/morality-as-autonomy doesn't take place there, but in the subsection entitled "Of the interest that attaches to the ideas of morality." Since my main objective in this paper is to present an interpretation of such deduction, I concentrate exclusively in the passages of *Groundwork III* where it occurs. Let me just mention that some authors, most notably Christine Korsgaard (1996), attach an exaggerated importance to the claim quoted above, as if the essence of Kant's argument were contained in it. This is a mistake since, as Kant makes clear, the crucial element in the deduction of the moral law is the idea of the two standpoints or worlds (discussed at length below in the text), not the fact that rational beings must act under the idea of freedom.

appears both in the premises and in the conclusion of the argument.[20] But in a Kantian deduction it makes perfectly good sense to say that freedom points towards the very thing that allows us to deduce it (in the sense of legitimating it). In what follows I will explain how this can be.

Recall that the positive concept of freedom is autonomy, that is, the capacity of the will to give laws to itself independently of the laws of the natural world. So the positive concept of freedom implies the suggestion that the will inhabits a world that is *not* the world of experience since, as Kant explains at length in the first *Critique*, experience only makes sense as a thoroughgoing connection of phenomena according to natural (causal) laws. This is why Kant says that the positive concept of freedom "provides" us with or "points us" towards the third thing linking rational will and autonomy, "which cannot, as in the case of physical causes, be the nature of the world of sense" (*G* 4, 447). As Kant will say four pages later, the world that the rational will inhabits according to the positive concept of freedom is the world of the understanding or the intelligible world. This is, in effect, the mysterious third thing that connects the concepts of a rational will and an autonomous will, and from which we can deduce the concept of freedom.

Thus, the idea is that the positive concept of freedom suggests *where* or *what* we have to look for to discover the connection between rationality and autonomy (this is the part just explained) and, in turn, the third thing in which the connection takes places—the intelligible world—can be used to show why we are *justified* in attributing ourselves an autonomous will, that is, a will that is free in the positive sense—this I explain in what follows. But before doing so, let us consider briefly the famous problem of the circle.

2.4 Interlude: the circle

We saw above that freedom plays the key role in linking the concepts of rational will and autonomy, although, according to Kant, freedom is not itself the linkage between them but only what points to the linkage. The problem of the circle (*G* 4, 450) appears because we have

[20] Ameriks (2003, 161, fn. 2) misses this point when he claims that in *Groundwork III* Kant attempts to provide a "strict deduction", by which he understands a "'linear' argument intended to be logically sound." However, Kant's argument cannot be "linear", precisely because the concept of freedom appears both in the premises and the conclusion.

to answer the question *why* we can attribute to ourselves freedom of the will in the negative sense, i.e., as independence of empirical causes. Kant's answer, as we will see below, is that in the faculty of reason, and more specifically in the production of ideas, we find in ourselves a capacity that is completely independent of the world of sense and that authorizes us to think ourselves as members of a different world (the intelligible world)—without ceasing to be, at the same time, members of the sensible world. But before presenting this argument, Kant addresses the worry that the only reason we might have for attributing freedom of the will to ourselves is that we are *already* committed to the moral law and, knowing that we can only do what it commands if we are free, we proceed to attribute freedom to ourselves for the law's sake. If this were the case, however, the whole strategy of *Groundwork III*—deducing the moral law from freedom—would be viciously circular.

Since my main purpose is to explain Kant's positive argument for the deduction of the moral law, I am going to skip a detailed exegesis of this problem.[21] I just want to point out that the fact that Kant detects here a potential threat to his strategy shows that the interpretation offered in the previous two subsections is in the right track: the concept of freedom cannot itself be what serves as the link between the concepts of rational will and autonomy since, as Kant explains, "freedom and the will's own legislation are both autonomy, and hence reciprocal concepts; but precisely because of this one of them cannot be used to explicate the other or to state its *ground*" (*G* 4, 450, emphasis added). But the fact that freedom cannot state the ground of the will's autonomy does not mean, as we saw in the previous subsection, that it cannot point us in the direction where such ground can be found. And, in turn, we can deduce freedom itself from this ground. This is what Kant turns to next.

2.5 The deductions, step three: the production of ideas and the spontaneity of reason

If the problem of the circle comes from the worry that the only reason we might have for attributing freedom of the will to ourselves is the authority we recognize in the moral law, its solution must come from discovering a reason for thinking ourselves as free that is *completely*

[21] For such a detailed exegesis, see McCarthy (1985).

independent of moral considerations.[22] This reason would be, at the same time, what explains or grounds the autonomy of the will. Interestingly, the very concept of negative freedom as independence from empirical causes suggests what this independent reason might look like: if when we think ourselves as free we locate ourselves in a different standpoint from which we think ourselves as phenomena, we could try to find out a non-moral reason that legitimates, first, the distinction between the two standpoints—sensible and intelligible—and, second, our attribution of membership in the intelligible one. If we succeed, then we would have dispelled the worry that our only reason for locating ourselves in the latter (or, what comes to the same thing, for thinking ourselves as free) is the authority of the moral law. Since for Kant moral law and practical reason are inextricable bound together, it is natural that in his argument for the two standpoints he makes use of considerations that have to do exclusively with *theoretical* reason.

Let us start with the question of why we can make the distinction between two standpoints or "worlds." Here Kant presents what can be taken as a very crude version of his argument in the Transcendental Aesthetic of the first *Critique* for the ideality of space and time and for the status of the objects of the senses as mere appearances.[23] He claims that even "the commonest understanding" (*G* 4, 450) can arrive at the conclusion that all representations in respect to which we are passive, i.e., those that come from the senses, "enable us to cognize objects only as they affect us, while what they may be in themselves remain unknown to us" (*G* 4, 451). From this conclusion we can draw the distinction between appearances and things in themselves and, furthermore, between the

[22] Tenenbaum (2012) claims that it is implausible to interpret Kant's argument in *Groundwork III* as an effort to locate non-moral reasons to think ourselves as free. Instead, he argues that already in this work Kant recognized that the only reasons we have for attributing freedom of the will to ourselves are moral ones. One of the main problems with Tenenbaum's contention is, precisely, that it makes unintelligible the whole problem of the circle and Kant's solution to it.

[23] Ameriks (1982, 215; 2003, 180) doubts that the argument presented in the *Groundwork* for the two standpoints can sensibly be taken even as a very crude summary of the arguments of the first *Critique* in favor of transcendental idealism. I hope that my interpretation will show that these doubts are unfounded.

world of sense and the world of understanding. This last distinction clearly exceeds what is concluded in the Transcendental Aesthetic; rather, it can be seen as deriving from further arguments that appear in the Resolution of the Third Antinomy to the effect that appearances themselves "must have grounds that are not appearances"—grounds which Kant identifies with the "intelligible cause" (*KrV* A 537/B 565) or with the "transcendental object" (*KrV* A 539/B 567).

Kant's next move confirms that he is employing arguments encountered in the Resolution of the Third Antinomy. He affirms that the same distinction between appearance and thing in itself must be applied to the self, since a human being cognizes herself only as she appears to herself through inner sense and consequently must presuppose "something else lying as its foundation, namely his I, such as it may be in itself" (*G* 4, 451), *and* (this is the part lifted from the Resolution) then Kant claims that this distinction between phenomenal and noumenal self corresponds, respectively, to a distinction between the faculties of sensibility and reason (*KrV* A 546-7/B 574-5). The further step Kant takes in the *Groundwork*, and the one that addresses the second question presented above (i.e., why we can adopt the intelligible standpoint) is to link explicitly the distinction sensibility/reason with the distinction between the world of sense and the (as he now calls it) the "intellectual world" (*G* 4, 451). This is the crucial step in the argument, since what Kant has to show now, without appealing to moral/practical premises, is that *we* inhabit (at least partially) the intellectual world. The fact that we are conscious of our possession of the faculty of reason is going to be the cornerstone for the conclusion that we do inhabit it.[24]

In effect, Kant proceeds by asserting that the human being does find in himself "a capacity by which he is distinguished from all other things, even from himself, in so far as he is affected by objects, and that is *reason*" (*G* 4, 452). As a further confirmation that Kant is appealing here to the Dialectic of the first *Critique*, he goes on to specify that it is in the production of ideas that reason "shows a spontaneity so pure that thereby he [the human being] goes far beyond anything that sensibility can ever afford him" (idem).[25] From here he moves quickly

[24] This crucial step is going to be the target of my two criticisms of Kant's argument in section 3 below.

[25] In subsection 3.1 below I discuss in some detail Kant's conception of ideas. Importantly, notice that Kant never says that he is thinking here about

to the conclusion that rational beings are entitled to view themselves as inhabiting the intelligible world; but if, as in the case of human beings, a rational being has sensibility in addition to reason, she has to regard herself "from the side of its lower powers" (*G* 4, 452) as at the same time belonging to the world of sense.

The crucial point here is that Kant thinks that the mere consciousness of the activity of reason in the production of ideas entitles us to make the following chain of inferences: if we are conscious of such activity, then we can consider ourselves as members of the intelligible world; hence as independent of the world of sense; hence as independent of natural laws; hence as free in the negative sense and, consequently, in the positive sense as well—which is, as we saw above, autonomy. In sum, starting from the following three premises drawn from theoretical philosophy: 1) the distinction between appearances and things in themselves as it is argued for in the Transcendental Aesthetic; 2) the further identification of the thing in itself with the intelligible cause of appearances in the Resolution of the Third Antinomy as the basis for the distinction between the worlds of sensibility and understanding;[26] and, finally, 3) the capacity of reason to produce ideas as described in the Dialectic, Kant is able to present a deduction of the concept of freedom that is not, at least ostensibly, supported by moral (or even

moral ideas; rather, he simply says that reason demonstrates its spontaneity in the production of ideas, period. So Tenenbaum's (2012, 583-585) argument that in this passage Kant is referring exclusively to practical/moral ideas—specifically to the idea of freedom as self-determination—is clearly unfounded. See also next footnote.

[26] There is, of course, the natural temptation to think that Kant's argument in the Resolution of the Third Antinomy is, at least partially, a moral argument, since one of his examples of the compatibility between freedom and causal necessity is precisely the human capacity to act based on moral imperatives, the results of which are appearances governed by causal laws (*KrV* A 547/B 575 and ff.). But this temptation must be resisted, since in the final paragraph of the Resolution Kant indicates that the purpose of his discussion has not been to prove the reality, or even the possibility, of freedom, but only to show that "this antinomy rests on a mere illusion, and that nature at least **do not conflict with** causality through freedom" (*KrV* A 558/B 586, boldface in the original). Hence, Kant doesn't provide a deduction of freedom from moral premises in the Resolution.

practical) considerations at all.[27] Given Henrich's definition of a Kantian deduction discussed in section 1, there is a clear sense in which this argument counts as a deduction (and a theoretically-grounded one)[28] of freedom: the origin of this concept is traced back to the faculties of the subject with the aim of legitimating its use.

If the deduction of freedom proceeds without making use of moral considerations, then the risk of a vicious circle is avoided and Kant can use the deduction to account for the synthetic *a priori* practical proposition, i.e., the moral law. Recall that in my reading this proposition is "a rational will necessarily is autonomous" and so the two terms that had to be linked are "rational will" and "autonomy". The deduction of freedom shows us how to do it, for it shows that a rational being, on account of her capacity to produce ideas, is entitled to regard herself as a member of the intelligible world and, consequently, as independent of the world of sense and its laws. But this just means that she can regard herself as free; and, given the further premise that a free will,[29] as a causality, cannot be lawless, it follows that the laws to which such a will is subject are the laws of the intelligible world, i.e., the laws of reason. So the free will, as the will of a being endowed with reason, obeys only

---

[27] Henrich (1998) questions this point. See footnote 29 for more details.

[28] It is crucial to bear in mind that a theoretically-grounded *deduction* of freedom is not the same as a theoretical *explanation* of the possibility of freedom, an explanation Kant considers to be impossible (*G* 4: 458-459). Ameriks (2003, 163) also emphasizes this point. See footnote 42 for more on the distinction between explanation and deduction.

[29] This subtle transition from the idea that a rational *being* is free to the idea that her *will* is free is the point where Henrich (1998, 336-337) suggests that Kant is introducing a premise that comes from practical, not theoretical, philosophy, since he claims that, according to Kant himself, the idea of a rational being *without* a will is not incoherent and, consequently, Kant can only appeal here to the fact that *we* are conscious of having a will—which, given Kant's robust definition of the will as the capacity to determine oneself to action as an intelligence, "*hence according to laws of reason*, independently of natural instincts" (*G* 4, 459, italics added), is already a practical (and moreover, moral) premise. I think Henrich has an important point here, although Kant could attempt to deflect it by appealing to the minimal definition of the will that appears at the beginning of Groundwork III as a "kind of causality" of rational beings (G 4, 446), instead of to the more robust one quoted above. In any case, my criticism of Kant's argument turns on a different point. See section 3.

the laws it gives to itself, which means that it is necessarily autonomous. In this way, the intelligible world fulfills its promised function as the third thing in which the concepts of rational will and autonomy come together and from where the concept of freedom can be deduced and, with it, the moral law.[30]

2.6 The deductions, step four: the deduction of the categorical imperative

Once Kant has provided a deduction of freedom, and with it a deduction of the moral law as autonomy, it can be hard to see what is missing.[31] What is missing, I claim, is a deduction of the moral law *as an imperative*.[32] In *Groundwork II* Kant distinguishes between an objective law of practical reason and an imperative (*G* 4, 412-413): an imperative is the form that an objective law takes when addressed to a rational

[30] This last step relies on what Allison (1990, Ch. 11) calls the "Reciprocity Thesis", that is, the idea that freedom and the moral law are reciprocal concepts. See footnote 32 for more details.

[31] For instance, Paton (1965, 247) explicitly claims that no work remains to be done by Kant after the (attempted) deduction of the moral law is concluded. Allison (1990, 227) agrees.

[32] Does this claim commit me to the position that in *Groundwork III* Kant offers a *triple* deduction—of freedom, of the moral law, and of the categorical imperative? The answer is no. The reason is that the deduction of freedom that I reconstructed in 2.5 above as "step three" can be seen as a deduction of the moral law at the same time, since at the beginning of *Groundwork III* Kant had explained that "if freedom of the will is presupposed, morality along with its principle follows from it, *by mere analysis of its concept*" (*G* 4, 447, emphasis added). Hence, once we have deduced (or legitimated) freedom of the will, we just have to do some conceptual analysis to realize that we have already deduced the moral law—so what we have here is just one deduction, not two. On the other hand, as I explain in the body of the text, the legitimation of the moral law as an *imperative* requires a further argument, which Kant provides in a separate subsection (entitled "How is a categorical imperative possible?"). Thus, although Kant does not make the distinction between the moral law—addressed to rational beings *simpliciter*—and the categorical imperative—as a command addressed to beings that combine rationality and sensibility—explicit in *Groundwork III*, the deduction of freedom/morality-as-autonomy and the deduction of the categorical imperative are clearly two separate arguments (again, I give my reason for this below in the text). In sum, I maintain my thesis that Kant performs a *double* deduction in *Groundwork III*: of freedom/morality-as-autonomy and of the categorical imperative.

will that is also subject to the influence of sensibility. The last step in the argument of *Groundwork III* has to show that the moral law takes an *imperatival* form for human beings, i.e., that for us it expresses itself as a categorical imperative. By doing so, the argument will legitimate the latter's pretension to be normative for us—in other words, it will provide a deduction of the imperative.

This is what happens in the subsection entitled "How is a categorical imperative possible?" (*G* 4, 453). The argument starts by restating the conclusion that a human being can consider herself from two standpoints; immediately afterwards Kant emphasizes that, from the standpoint of the intelligible world, all human actions occur according to rational laws (i.e., laws of autonomy), whereas from the standpoint of the sensible world they take place following natural laws (what Kant calls "heteronomy of nature" [*G* 4, 453]). The key addition to the argument—and what proves that this is a *different* argument from the one that deduces freedom/morality-as-autonomy—is the premise that "the world of understanding contains the ground of the world of sense, *and hence also of its laws*" (idem, emphasis added).

In the previous stage of the argument we were told that the "commonest understanding" assumes that behind the appearances stand things in themselves (*G* 4, 451), and we were also told that a human being can recognize two different sets of "laws for the use of [her] powers, and consequently for all [her] actions" (*G* 4, 452) corresponding to the two standpoints. What we were *not* told, however, is that the intelligible world (the realm of things in themselves) gives laws to the sensible world (the realm of appearances). Given this additional premise, Kant is in a position to infer that the pure will of a rational being—a will "which belongs wholly to the world of understanding" (*G* 4, 453)—is legislative for her will "affected by sensuous desires" (*G* 4, 454), that is, for the will that belongs to the sensible world.

This relation between a sensibly affected rational will and objective practical laws is called by Kant "necessitation" (*G* 4, 413). Since it doesn't inevitably occurs that the former conforms to objective laws, this kind of will *ought* to conform to them or, in other words, is necessitated to do so. The representation of necessitation is called by Kant a "command," and the formula of a command is an "imperative" (idem). Thus, the relation between a lawgiving will and a subordinate will explains how a categorical imperative is possible.

I want to emphasize two points about the deduction of the categorical imperative. First, I claim that this counts as (at least an attempt to perform) a deduction in the specific Kantian sense since, as in the case of freedom, the legitimation of the concept (in this case, command) proceeds again by searching its origin in the subject—in this case, in the relation between the subject's pure and empirical wills. Moreover, in *Groundwork II* Kant explains that the question of the *possibility* of the imperatives (hypothetical and categorical) is a question about "the necessitation of the will that the imperative expresses in its task" (*G* 4, 417). And, as we just saw, the last argument addresses precisely the issue of how to understand and legitimate the necessitation expressed by the categorical imperative; in this sense, the argument accounts for the possibility of the latter, and thus it should be deemed a deduction.[33]

Second, I wish to present one additional piece of evidence to the effect that the deduction of the categorical imperative is different from the deduction of freedom/morality-as-autonomy. The evidence is that the synthetic *a priori* proposition that Kant claims to be grounding in the subsection entitled "How is a categorical imperative possible?" is different from the one that he is grounding in the one entitled "Of the interest that attaches to the ideas of morality" (where the deduction of freedom/morality-as-autonomy takes place). In the latter case, and as I explained at length above, it is clear that the two terms that Kant wants to link are "rational will" and "autonomy," and so the synthetic *a priori* proposition that needs to be explained is "a rational will is necessarily autonomous." By contrast, in the former case Kant suggests that the two terms that have to be linked are "will affected by sensuous desires" and "will belonging to the world of the understanding" (*G* 4, 454); hence, the synthetic *a priori* proposition that needs to be deduced here is something like "the empirical will ought to conform to the laws of the intelligible will." Thus, given that freedom/morality-as-autonomy and the categorical imperative are expressed as two different synthetic *a*

[33] As I mentioned in section 1, Henrich claims that the distinctively Kantian "How is it possible?" question is always answered through a deduction (1989, 35). Rauscher (2009, 223) mistakenly claims that the "How is it possible?" question regarding the categorical imperative can only be partially answered, given that we cannot explain how it is that our noumenal self affects our empirical will. This claim betrays a confusion between the very different tasks of explanation and deduction. See footnotes 28 and 42 for more on this distinction.

priori propositions, Kant needs two separate deductions to demonstrate their validity.[34]

3 Two problems with the deduction of freedom/morality-as-autonomy in *Groundwork III*

The above reconstruction of Kant's deductions in *Groundwork III* has, I think, interest in itself, since it allows us to see in detail Kant's strategy for presenting a theoretically-grounded deduction of the central practical concepts. Even more interesting, however, is to understand the problems that Kant saw in these deductions and that convinced him to give them up in the *Critique of Practical Reason*.[35] In effect, not only does Kant claim there that "the objective reality of the moral law cannot be proved by any deduction" and that the moral law "itself has no need of justifying grounds" (*KprV* 5, 47), but he also completely reverses the order of explanation. Surprisingly, he now claims that "something different and quite paradoxical takes the place of this *vainly sought deduction* of the moral principle" (idem, italics added),[36] namely, that the latter is what makes possible the deduction of freedom. In this section I will investigate two possible sources of Kant's discomfort with his arguments in *Groundwork III*, both of which have to do with the crucial step of securing a non-moral reason for thinking ourselves as members

[34] This shows that Paton (1965, 247), Allison (1990, 227), Tenenbaum (2012, 580-581), McCarthy (1979), and Timmermann (2010), all of whom defend the idea that only a single deduction (either of freedom/morality-as-autonomy or the categorical imperative) takes place in *Groundwork III*, are mistaken.

[35] In this section I concentrate only in the deduction of freedom/morality-as-autonomy. However, since the deduction of the categorical imperative depends on the success of the former, any problems that afflict the deduction of freedom are also relevant for the deduction of the imperative. See Allison (1990, 225-226) for a direct criticism of the latter.

[36] I think that Kant's wording of these passages pretty much suffices to prove that those interpreters who insist that there is no great reversal between *Groundwork III* and the second *Critique* are wrong. For in the former work Kant explicitly refers to what he has done as a "deduction of the supreme principle of morality" (*G* 4, 463), whereas in the latter, as we just saw, he talks of the "vainly sought deduction of the moral principle." One of Tenenbaum's (2012, 557) arguments for denying the great reversal is that Kant never explicitly recanted the argument offered in *Groundwork III*; however, comparing these two passages shows that such recantation indeed took place.

of the intelligible world. First, I will take issue with Kant's claim that the spontaneity reason exhibits in the production of speculative ideas suffices for attributing freedom in the negative sense to the *will* of a rational being. Second, I will question whether, given the restrictions of Kant's critical philosophy, it is legitimate to say, as Kant does in *Groundwork III*, that the consciousness of this spontaneity legitimates the subject in conceiving herself as a noumenon and so as "belonging to the *intellectual world*" (*G* 4, 451).[37] Before all that, however, I will present a brief overview of Kant's theory of ideas.

3.1 Kantian ideas: speculative and practical

In the first book of the Transcendental Dialectic in the *Critique of Pure Reason* Kant presents his theory of ideas as the concepts generated

[37] Allison (1990, 227-229) suggests that the central problem in the deduction lies elsewhere, namely, in an ambiguity in two central concepts: intelligible world and will. Timmermann (2010, 78-79) argues that the problem is rather that we lack an intuition of the intelligible world and, he adds, for Kant an intuition is necessary to confirm the validity of any deduction (this is clear in the case of the deduction of the categories). Unfortunately, I don't have the space to discuss the merits of these proposals in detail, so I will just say a quick word about them. Allison claims that Kant's argument trades upon an ambiguity in the concept of the intelligible world because all that we can conclude from the spontaneity of reason is that we belong to the intelligible world *negatively* conceived, i.e., as that which excludes everything sensible; however, Allison continues, the conception of the intelligible world Kant needs for the success of his deduction is the *positive* conception according to which the intelligible world is nothing other than the Kingdom of Ends. This objection to Kant's argument is, I think, implausible on its face, at least for two reasons: first, it is simply false that Kant needs the positive conception in order for the deduction to go through; at this point in the argument, all he needs is a standpoint from which we can think ourselves as free from all sensible influences, and for this the negative conception of the intelligible world is clearly sufficient. Second, it would be completely circular for Kant to deduce the moral law from our belonging to the intelligible world conceived as the Kingdom of Ends, since the latter is a moral idea. Now concerning Timmermann's objection, the problem is that it mistakenly assumes that deductions of theoretical and practical concepts have to be parallel in every respect. But they don't; rather, what makes them deductions is the strategy of legitimating a concept by tracing its origin in the subject's rational capacities. So a deduction of a practical concept doesn't need, as Timmermann assumes, an intuition supporting it.

by the faculty of reason. Kant defines an idea of reason as follows: "A concept made up of notions [i.e., pure concepts of the understanding], which goes beyond the possibility of experience, is an **idea** or concept of reason" (*KrV* A 320/B 377). To fully understand this definition, we must go back to Kant's discussion of reason in the Introduction to the Dialectic. Kant defines reason there as the faculty of principles (*KrV* A 299/B 356) and defines a cognition from principles as "that cognition in which I cognize the particular in the universal through concepts" (*KrV* A 300/B 357). The characteristic activity of reason in its logical use is precisely to seek "the universal condition of its judgment (its conclusion)" (*KrV* A 307/B 364), and the paradigmatic form of this type of cognition is the syllogism. Now since the universal condition (or major premise) of a syllogism can itself be subsumed under another universal condition, reason is compelled to seek the latter "by means of a prosyllogism" (idem); but, since the universal condition in the prosyllogism can once more be subsumed under a more general rule, reason pursues again this more general rule. Given that reason does not stop its inquiry until it reaches the unconditioned, Kant concludes that "the proper principle of reason in general (in its logical use) is to find the unconditioned for conditioned cognitions of the understanding, with which its unity will be completed" (idem).

Following the example laid down in the Transcendental Analytic concerning the understanding, in the first book of the Dialectic Kant suggests that the logical use of reason contains the key for discovering the pure concepts of reason or transcendental ideas (*KrV* A 321/B 377-8). In particular, the logical function of reason—by which reason seeks more and more universal conditions under which to subsume the major premises of syllogisms—begets the transcendental concept of "the totality of conditions to a given conditioned thing" (*KrV* A 322/B 379). This is, in effect, the master concept through which all other concepts or ideas of reason can be explained. Kant elaborates as follows:

> Now since the **unconditioned** alone makes possible the totality of conditions, and conversely the totality of conditions is always itself unconditioned, *a pure concept of reason in general can be explained through the concept of the unconditioned*, insofar as it contains a ground of synthesis for what is conditioned (*KrV* A 322/B 379, italics added, boldface in the original).

But the concept of the unconditioned is necessarily transcendent, that is, it necessarily goes beyond the limits of experience, since "the absolute totality of conditions is not a concept that is usable in an experience, because no experience is unconditioned" (*KrV* A 326/B 383). So now we can see why Kant defines a transcendental idea as a concept which goes beyond the possibility of experience: such an idea is based on the concept of the unconditioned, and so nothing given in experience can correspond to it.

Later on Kant distinguishes between an idea of reason in the speculative and in the practical use of this faculty. It is hard to give a non-circular definition of a speculative or a practical idea, but the following is one way of grasping the difference. On the one hand, a speculative idea is one that behaves as if it were a concept of the understanding in the sense that it *purports* to apply *a priori* to experience by demanding the totality of conditions for a given conditioned.[38] In effect, reason employs the principle: "If the conditioned is given, then the whole sum of conditions, and hence the absolutely unconditioned, is also given" (*KrV* A 409/B 436) to ground its demand that for any series of given conditions (e.g., a causal chain) the unconditioned must be given as well (e.g., a first cause)—a demand that persists regardless of whether the sum total of the conditions is even possible in experience. The role Kant assigns to speculative ideas in the architectonic of reason is to "serve the understanding as a canon for its extended and self-consistent use, through which it cognizes no more objects than it would cognize through its concepts, yet in this cognition it will be guided better and further" (*KrV* A 329/B 385). In other words, although speculative ideas cannot expand knowledge because what they demand of experience (the unconditioned) cannot be given in it, they can still assist the understanding in its own tasks by presenting themselves as regulative principles for the empirical use of the latter (*KrV* A 509-510/B 537-538; also *KrV* A 516/B 544).

[38] "[R]eason really cannot generate any concept at all, but can at most only **free** a **concept of the understanding** from the unavoidable limitations of a possible experience, and thus seek to extend it beyond the boundaries of the empirical, *though still in connection with it*" (*KrV* A 409/B 435, emphasis added). See also: "[T]ranscendental ideas will really be nothing except categories extended to the unconditioned" (*KrV* A 409/B 436).

On the other hand, a practical idea is a guideline for action. Given that reason in its practical use is concerned not with knowledge but with "execution according to rules" (*KrV* A 328/B 385), Kant claims that "an idea of practical reason can always be actually given *in concreto*, though only in part; *indeed, it is the indispensable condition of every practical use of reason*" (idem, emphasis added). Two things are important here: first, practical ideas, unlike speculative ones, are immanent, in the sense that their objects can be given—though always only partially—in experience through the exertion of practical reason itself.[39] Second, it is not as if reason was practical independently of the ideas but, as the italicized passage above explicitly claims, reason can only be practical *through* ideas, that is, the only way in which reason can influence conduct is by way of them.[40]

The distinction between speculative and practical ideas corresponds to the distinction between two types of spontaneity that, although Kant does not explicitly present, clearly follows from the differences between the two kinds of ideas. On the one hand, theoretical or speculative spontaneity is the capacity of reason to extend the categories beyond experience—an extension that, although cannot yield *knowledge*, does afford reason the possibility of *thinking* beyond the limits of experience. In particular, it affords reason the possibility of thinking the unconditioned. On the other hand, practical spontaneity is the capacity of reason to determine conduct based exclusively on its ideas, independently of the influences of sensibility and so independently of natural causality. Hence, the spontaneity of practical reason is nothing less than transcendental freedom, which Kant defines as "the faculty

[39] A legitimate question is how practical ideas so described fit the description Kant gives of ideas in general as emerging from the concept of the unconditioned and so as transcending the limits of experience. Two points are relevant here: first, practical ideas can only be given *partially* in experience, that is, it is impossible that empirical *examples*, say, of virtue or of the just State, correspond to the *idea* of perfect virtue or to the idea of the perfectly just State (*KrV* A 315-7/B 371-4). Second, and closely related, practical ideas can be said in this sense to be based on the concept of the unconditioned, since, as Kant writes, their execution always occur "under the influence of the concept of an absolute completeness" (*KrV* A 328/B 385).

[40] See *G* 4, 427: "because if *reason all by itself* determines conduct … it must necessarily do this *a priori*." The only way in which reason can determine conduct *a priori* is precisely through its ideas.

of beginning a state **from itself**, the causality of which does not in turn stand under another cause determining it in time in accordance with the law of nature" (*KrV* A 533/B 561, boldface in the original).[41] In short, practical spontaneity is the capacity of reason to initiate an action *de novo* guided by ideas and independently of the causal law of nature.

3.2 Speculative ideas and freedom of the will

Now we can finally pose the question relevant to the deduction of freedom/morality-as-autonomy in *Groundwork III*: Is the spontaneity of reason Kant appeals to in *Groundwork III*, and which plays the central role in attributing negative freedom to a rational being, the spontaneity of theoretical or practical reason? In light of my reconstruction of Kant's deduction—especially in relation to the problem of the circle—it is easy to see that this is a rhetorical question: Kant *has* to appeal only to the spontaneity of reason in the production of *speculative* ideas, because otherwise his argument would be trapped in the circle it was meant to avoid. To see why, consider the following: we saw above that practical ideas are what *make* reason practical, that is, it is only through such ideas that reason can determine by itself the conduct of a rational being. If this is so, then the spontaneity that reason shows in its *practical* use is the spontaneity of determining conduct through ideas, that is, the capacity to determine conduct *a priori* independently of any influence from sensibility. But this is plainly the negative conception of freedom that Kant presents at the beginning of *Groundwork III*, which the argument from the spontaneity of reason was supposed to legitimate. Precisely because of this, such negative conception cannot be assumed at the outset of the argument.

The only alternative is, then, that the spontaneity Kant is appealing to in the argument for negative freedom is the spontaneity of reason in the production of speculative ideas. This matches well with what I claimed above is Kant's solution to the problem of the circle, namely, to find a non-moral reason for attributing freedom to ourselves. What I have added here is the explanation why "non-moral" has to be interpreted more broadly as "non-practical," namely because the spontaneity of practical reason is too closely connected to morality and

[41] See *KrV* A 533/B 561 for Kant's identification of transcendental freedom with practical spontaneity.

so cannot be used as a non-moral premise. The pressing question we must now consider is whether the spontaneity of theoretical reason can yield the conclusion Kant wants to derive from it, namely, that the *will* of a rational being is negatively free.

The answer I suggest is this: now that we have in clear sight the fact that a will that is negatively free must necessarily be characterized as spontaneous, we can see the extremely dire prospects of any argument that attempts to derive freedom of the will from the spontaneity of theoretical reason, for, in effect, such an argument attempts to derive the *practical* spontaneity of reason from the *speculative* one. And, given the two very different roles that Kant assigns to speculative and practical ideas, and so the two very different senses in which reason can be spontaneous, the enterprise of deriving one kind of spontaneity from the other seems hopeless. We can put the same point in an even stronger form: since Kant claims that reason can be practical only because it exhibits spontaneity through practical ideas, the project of deriving practical spontaneity from speculative spontaneity is equivalent to attempting to show, from an exclusively speculative standpoint, that reason is indeed practical.[42] Given that Kant's argument in *Groundwork III* has at bottom precisely this form, it is not wonder that eventually Kant himself noticed its inherent problems and decided to abandon it.

In sum, in appealing to the spontaneity of reason in the production of ideas to prove that the will of a rational being is negatively free Kant

[42] To avoid misunderstandings, it is crucial to see that what is at stake here is *not* to *explain* how pure reason can be practical. In *Groundwork III*, in the subsection entitled "On the extreme boundary of all practical philosophy" (which comes after the deductions of freedom and of the categorical imperative), Kant claims that such an explanation is not an option: "But reason would overstep all its bounds if it undertook to *explain* **how** pure reason can be practical, *which would be one and the same task entirely* [emphasis added] as to explain *how freedom is possible*" (*G* 4, 458-459). Since Kant claims here that the possibility of freedom cannot be explained, just after he has provided a *deduction* of freedom, it is patent that explanation and deduction must be two different tasks: one can deduce freedom, but one cannot explain it. The point I am making in the body of the text concerns deduction, not explanation, and what I am claiming is that attempting to deduce (in the Kantian sense of legitimating) freedom of the will from speculative spontaneity is exactly the same as trying to deduce (or legitimate the use of) practical reason from theoretical reason. And I suggest that such deduction is hopeless by Kant's own lights.

faced the following dilemma: on the one hand, if the spontaneity in question is practical spontaneity, then he assumes from the outset what the argument was supposed to show, namely, that the will of a rational being is capable of determining conduct independently of sensibility and hence that it is free. On the other hand, if the spontaneity in question is speculative spontaneity, then it is entirely unclear how the fact that reason shows spontaneity in the production of speculative ideas has any direct relevance to the question of whether practical reason is capable of the very different spontaneity that it exhibits when determines conduct *a priori*. The former is the spontaneity to *think* beyond the limits of experience, whereas the latter is the spontaneity to *act* without being empirically influenced. Within the confines of the Kantian system, there simply seems to be no way to get from one to the other.[43]

3.3 Spontaneity of reason and the self as noumenon

The second crucial problem that afflicts Kant's deduction of freedom/morality-as-autonomy is that it relies on the following inference: given the capacity to produce speculative ideas, the subject is entitled to conceive herself as noumenon and so as free. The question we should ask is whether this inference is a legitimate one for Kant to make. The answer I shall give, based on Kant's own mature theory of the self, is negative.

Before offering the reason for this answer, let me present two passages where it is patent that Kant is making the aforementioned inference. First, in the course of the deduction of freedom in *Groundwork III*, and after claiming that the subject must assume that behind his self as an appearance lies "his I, such as it may be in itself," Kant claims that "with regard to what there may be of pure activity in him (what reaches consciousness not by affection of the senses, but immediately)

[43] An obvious objection is that if Kant's argument for deducing freedom from theoretical premises is so deeply mistaken, it is very uncharitable on my part to attribute it to him. A quick response is that, as Henrich (1994) shows, Kant indeed attempted for a very long time (from the mid-1760s onwards and, I have argued, until 1785, the year of the *Groundwork*'s publication), and from many different routes, to achieve a deduction of freedom exclusively from theoretical premises. The main innovation of *Groundwork III* was to try to deduce freedom from speculative spontaneity not directly, but through the mediation of the intelligible world. (On this point, see Allison [1990, 223].)

[the subject must count himself] as belonging to the *intellectual world*" (*G* 4, 451). Two paragraphs later, Kant makes it clear that by "pure activity" he means the capacity of reason to produce ideas, which, as I argued at length above, must be interpreted as the capacity to produce *speculative* ideas.

Second, in the Resolution of the Third Antinomy in the *Critique of Pure Reason*, and after explaining the difference between the empirical and intelligible character of an acting cause (*KrV* A 539/B 567) and arguing that both can coexist in the production of the same actions (*KrV* A 541/B 569), Kant offers as an example of the possibility of this coexistence the double character of human beings as appearances and things in themselves. He writes:

> Yet the human being … knows himself also through pure apperception, and indeed in actions and inner determinations which cannot be accounted at all among impressions of sense; he obviously is in one part phenomenon, but in another part, namely in regard to certain faculties, he is a merely intelligible object … We call these faculties understanding and reason; chiefly the latter is distinguished quite properly and preeminently from all empirically conditioned powers, since it considers its objects merely according to ideas and in accordance with them determines the understanding (*KrV* A 546-547/B 574-575).

Notice that in this passage Kant appeals not only to ideas (produced by reason) but also to pure apperception (a function of the understanding) to build his case that human beings are intelligible objects. But, to repeat the question posed above: Is Kant entitled to conclude from these cognitive operations—the production of ideas and pure apperception—that human beings exist (or must conceive themselves) as things in themselves?

The answer is clearly negative, and Kant himself explains why in two passages from the B edition of the first *Critique*. First, in the §25 of the deduction of the categories as it appears in the B edition, Kant explicitly claims that from pure apperception the subject cannot infer his existence as a noumenal self: "In the transcendental synthesis of the manifold of representations in general … hence in the synthetic original unity of apperception, I am conscious of myself not as I appear to myself, *nor **as***

I am in myself, but only **that** I am" (*KrV* B 157, italics added, boldface in the original). And then, in a footnote to the same section, Kant adds the following:

> I cannot *determine* my existence as that of a self-active being, rather I merely *represent* the spontaneity of my thought … and my existence always remains only sensibly determinable … Yet this spontaneity is the reason I call myself an **intelligence** (*KrV* B 158 n., italics added).

This passage is interesting for two reasons. First, in it Kant draws the distinction between *determining* one's existence as a self-active being (or noumenon) and *representing* the spontaneity of one's thought, a distinction that clearly implies that one cannot infer the former from the latter. Furthermore, since he claims that the subject can call himself an intelligence due to this spontaneity, this implies that, from the fact that one can *conceive* oneself as an intelligence, it does not follow that one can *determine* one's existence as a noumenon, which is precisely what Kant improperly does in the deduction of freedom in *Groundwork III*.[44] Second, the passage seems to imply that the spontaneity of reason does not suffice to determine our existence as self-active beings either, since, as Kant emphasizes, "my existence always remains only sensible determinable." If this is Kant's considered position, as I think it is, then it is not only the spontaneity of the understanding that is insufficient for determining our existence as noumena, but that of reason as well.

The second relevant passage is found in the General Remark that appears at the end of the Paralogisms in the B edition. In it, Kant insists on the limits of what can be inferred from the spontaneity of thought:[45]

> Thinking, taken in itself, is merely the logical function and hence *the sheer spontaneity of combining the manifold*

[44] For more on the distinction between determining and representing one's existence, see footnote 47 below.

[45] Ameriks (2003, 182) also emphasizes the relevance of the General Remark for explaining the great reversal; however, he doesn't take into account the difference between the spontaneity of the understanding and of reason and, as a consequence, he doesn't consider the obvious objection I refer to immediately in the body of the text.

> of a merely possible intuition … In this way I represent myself to myself *neither as I am* nor as I appear to myself, but rather I think myself only as I do every object in general from whose kind of intuition I abstract (*KrV* B 428-429, italics added).

Again, it is clear that in Kant's revised theory of the self[46] the spontaneity the subject exhibits in thinking provides no ground at all for inferring anything about his supersensible existence. An obvious objection is that the spontaneity Kant is referring to in this passage is the spontaneity of the understanding (since he talks about the spontaneity of combining the manifold of intuition), and in *Groundwork III* (*G* 4, 452) Kant explicitly says that it is the spontaneity of *reason*, not of the understanding, which justifies the subject's pretension to be a member of the intelligible world. However, once we recall the footnote in the §25 of the B deduction of the categories quoted above, it becomes clear that Kant's considered view is that, despite the different kinds of spontaneity exhibited by reason and the understanding, the mere spontaneity of thought cannot be employed—as he did in *Groundwork III*—to determine our existence as intelligible beings.

In sum, in these passages Kant discredits the inference from speculative spontaneity to the supersensible existence of the subject and so he discredits the argument he offers in *Groundwork III* for the deduction of freedom/morality-as-autonomy, whose ground is precisely that inference. It is noteworthy that in the *Critique of Practical Reason* (published after the revised edition of the first *Critique*) Kant not only abandons the strategy of grounding freedom and the moral law in the spontaneity of thought, but also claims that the immediate consciousness of the moral law's bindingness (the Fact of Reason) serves

[46] Revised because these passages are drawn from parts of the *Critique* that Kant completely rewrote for the second edition. Tenenbaum (2012, 557) claims that an important reason for thinking that there isn't a great reversal in the second *Critique* is that between 1785 and 1788 Kant's views concerning freedom, morality, and the relation between the sensible and the intelligible worlds underwent no major change. But the last contention is false regarding the relation between the sensible and the intelligible *self*: in the B edition of the first *Critique*, completed in 1787, Kant's theory of the self did change in important respects and, as I argue below, this change provides both historical and philosophical support for the thesis that a great reversal did occur.

as the cornerstone for the deduction of freedom and for the postulate of a supersensible existence. Interestingly, both of these points are already suggested in the General Remark of the B Paralogisms (*KrV* B 430-1).[47] Here, then, it is plausible to suggest that the philosophical and the historical arguments finally meet since, as Karl Ameriks claims, "The revising of the first *Critique* and the forsaking of the deduction of the *Foundations* take place simultaneously" (1982, 217). Hence, it is a legitimate speculation that what finally convinced Kant that the deduction of freedom—and with it the deduction of the moral law—in *Groundwork III* was wrong was precisely the impossibility, given the restrictions of his critical system, to rely on the inference from the spontaneity of thought to the noumenal existence of the subject.

Conclusion

In this paper I have provided an interpretation and evaluation of the argument in *Groundwork III*. I claimed that the argument is a double deduction—of freedom/morality-as-autonomy and of the categorical imperative—and explained why, based on Henrich's rendering of what a Kantian deduction is, the argument in that section is indeed a deduction (or at least an attempt thereof). Then I raised two criticisms of Kant's argument. First, that appealing to the spontaneity of reason for attributing freedom of the will is either circular or irrelevant. Second, that Kant's argument relies on an illegitimate inference from speculative spontaneity to the supersensible existence of the thinking subject as a thing in itself—an illegitimate move in light of Kant's mature theory of

[47] Regarding freedom: "But suppose there subsequently turned up – not in experience but in certain (not merely logical rules but) laws holding firm *a priori* and concerning our existence – the occasion for presupposing ourselves to be **legislative** fully *a priori* … then this would *disclose* a spontaneity through which our actuality is determinable without the need of conditions of empirical intuition" (*KrV* B 430, emphasis added). Regarding supersensible existence: "For through this admirably faculty [to be legislative fully *a priori*], which for the first time reveals to me the consciousness of the moral law, I would indeed have a principle for the *determination* of my existence that is purely intellectual" (*KrV* B 431, emphasis added). Notice how, in the latter passage, Kant talks about the *determination* of the subject's existence as an intelligible being through the moral law, and compare this with what Kant claims can be achieved through the mere spontaneity of thought in the footnote at B 158 quoted above.

the self. Finally, I indicated that in the General Remark that appears at the end of the B Paralogisms we can see foreshadowed Kant's doctrine of the Fact of Reason that substitutes the deduction of the moral law in the *Critique of Practical Reason*. This constitutes evidence for the conjecture that Kant abandoned the project of a theoretical deduction of the moral law due to his considered conception of what can (and cannot) be legitimately inferred from the subject's spontaneity in thinking.

The main conclusion of this investigation is that the deductions in *Groundwork III* present structural problems that, given Kant's own doctrines about the different functions of theoretical and practical reason, and the clear boundaries that his considered view of the self establishes between the consciousness of spontaneity and supersensible existence, strongly recommend abandoning the whole enterprise. This result is satisfactory, given that the point of departure of the investigation was Kant's explicit disavowal of the possibility of deducing the moral law in the second *Critique*.

Finally, I hope to have proven wrong those interpreters[48] who deny that the argument in *Groundwork III* is an attempt to perform a theoretically-grounded deduction of freedom (and the moral law) and who, as a consequence, deny that a great reversal occurs in the second *Critique*. And I also hope to have proven wrong those who insist that only a single deduction takes place in the former work.[49]

References

Allison, H. (1990). *Kant's Theory of Freedom*. Cambridge: Cambridge University Press.

Ameriks, K. (1982). *Kant's Theory of Mind*. Oxford: Oxford University Press.

—— (2003, originally published 1981). Kant's Deduction of Freedom and Morality. *Interpreting Kant's Critiques*. (161-192). Oxford: Oxford University Press.

[48] Paton (1965), Henrich (1998), Esteves (2012), Tenenbaum (2012) and, more tentatively, Rauscher (2009).

[49] Paton (1965, 247), Allison (1990, 227), Tenenbaum (2012, 580-581), McCarthy (1979), and Timmermann (2010).

Esteves, J. (2012). The Non-Circular Deduction of the Categorical Imperative in *Groundwork III*. *Kant in Brazil*. D. Perez & F. Rauscher (Eds.) (155-172). Rochester: University of Rochester Press.

Henrich, D. (1989). Kant's Notion of a Deduction and the Methodological Background of the First *Critique*. In *Kant's Transcendental Deductions*. E. Forster (Ed.) (29-46). Stanford, CA: Stanford University Press.

____ (1994, originally published in 1960). The Concept of Moral Insight and Kant's Doctrine of the Fact of Reason. *The Unity of Reason*. (55-88). Cambridge, MA: Harvard University Press.

____ (1998, originally published in 1975). The Deduction of the Moral Law: The Reasons for the Obscurity of the Final Section of Kant's *Groundwork*. In *Kant's Groundwork of the Metaphysics of Morals: Critical Essays*. P. Guyer (Ed.) (303-342). New York: Rowman & Littlefield.

Kant, I. (1997). *Critique of Practical Reason*. M. Gregor (Trans.) Cambridge: Cambridge University Press.

____ (2007). *Critique of Pure Reason*. P. Guyer & A. Wood (Trans.) Cambridge: Cambridge University Press.

____ (2012). *Groundwork of the Metaphysics of Morals*. M. Gregor & rev. J. Timmermann (Trans.) Cambridge: Cambridge University Press.

Korsgaard, C. (1996). Morality as Freedom. *Creating the Kingdom of Ends*. (159-187). Cambridge: Cambridge University Press.

McCarthy, M. (1979). Kant's Application of the Analytic/Synthetic Distinction to Imperatives. *Dialogue* 18, 373-391.

____ (1985). The Objection of Circularity in *Groundwork III*. *Kant-Studien* 76, 28-42.

Paton, H. (1965). *The Categorical Imperative*, London: Harper.

Rauscher, F. (2009). Freedom and Reason in *Groundwork III*. In *Kant's 'Groundwork of the Metaphysics of Morals': A Critical Guide*. J. Timmermann (Ed.) (203-223). Cambridge: Cambridge University Press.

Tenenbaum, S. (2012). The idea of Freedom and Moral Cognition in *Groundwork III*. *Philosophy and Phenomenological Research*, 84, 555-589.

Timmermann, J. (2010). Reversal or Retreat? Kant's Deductions of Freedom and Morality. In *Kant's 'Critique of Practical Reason': A Critical Guide*. A. Reath and J. Timmermann (Eds.) (73-89). Cambridge: Cambridge University Press.