http://dx.doi.org/10.15446/ideasyvalores.v67n166.62775

KANT ON FREE WILL AND THEORETICAL RATIONALITY*

LIBRE ALBEDRÍO Y RACIONALIDAD TEÓRICA EN KANT

DANIEL M. WOLT**
Universidade de São Paulo - São Paulo - Brazil

Article received on February 12, 2017; approved on July 27, 2017.

* This work was written in part with the aid of financial support from the São Paulo Research Foundation (Fapesp; Grant n.º 2016/05983-0).

** *woltdm@gmail.com*

How to cite this article:

MLA: Wolt, D. "Kant on Free Will and Theoretical Rationality." *Ideas y Valores* 67.166 (2018): 181-198.

APA: Wolt, D. (2018). Kant on Free Will and Theoretical Rationality. *Ideas y Valores*, 67 (166), 181-198.

CHICAGO: Daniel Wolt. "Kant on Free Will and Theoretical Rationality." *Ideas y Valores* 67, n.º 166 (2018): 181-198.

ABSTRACT

The focus of this essay is Kant's argument in the *Groundwork for the Metaphysics of Morals* (GMS) III that regarding oneself as rational implies regarding oneself as free. After setting out an interpretation of how the argument is meant to go (§§1-2), I argue that Kant fails to show that regarding oneself as free is incompatible with accepting universal causal determinism (§3). However, I suggest that the argument succeeds in showing that regarding oneself as rational is inconsistent with accepting universal causal determinism if one accepts a certain, plausible view of the explanation of events (§4).

Keywords: I. Kant, free will, *Groundwork for the metaphysics of morals*, theoretical rationality.

RESUMEN

El ensayo se enfoca en el argumento de Kant en la *Fundamentación de la metafísica de las costumbres* (GMS) III según el cual considerarse racional implica verse a uno mismo como libre. Se interpreta la forma en que debe entenderse el argumento (§§1-2) y se afirma que Kant no logra demostrar que considerarse libre es incompatible con la aceptación del determinismo casual universal (§3). No obstante, se sugiere que el argumento sí logra demostrar que considerarse a uno mismo como racional es incompatible con la aceptación del determinismo casual universal, si se acepta una cierta versión plausible de la explicación de los eventos (§4).

Palabras clave: I. Kant, libre albedrío, *Fundamentación de la metafísica de las costumbres*, racionalidad teórica.

The Argument in Context

Philosophers most often deal with the question of free will in connection with moral issues. For example, the question of whether determinism and free will are compatible is usually of interest to people who are concerned about the implications of determinism for moral responsibility: free will is a prerequisite for moral responsibility, the thought goes, so if determinism rules out free will and determinism is a fact, then none of us are morally responsible for our actions. It is unsurprising, then, that Kant's main interest in free will also has to do with questions pertaining to morality. What is surprising, though, is that Kant not only thinks that freedom of some kind must be presupposed in order to regard ourselves as morally responsible, but also that freedom must be presupposed in order to regard ourselves as theoretically rational (*i.e.* capable of rationality in the theoretical, as opposed to the practical, domain). But why exactly Kant thinks this is not entirely easy to determine. Here I will focus on the argument as it occurs in the *Groundwork for the Metaphysics of Morals*, though, of course, in attempting to explain that argument, I will make reference to what he says about the issue elsewhere as well.

Kant's goal in the third section of the *Groundwork* is to show that the categorical imperative applies to all rational agents. In order to do so, Kant needs to show that all rational agents must be regarded as free in some way, because in order for the categorical imperative to apply to an agent, that agent must be capable of acting or failing to act accordingly. Thus, Kant begins the first part of section III ("The Concept of Freedom is the Key to the Explanation of the Autonomy of the Will") with the analysis of the concept of freedom. He starts with a negative characterization: freedom is a property of the will "that it [the will] can be efficient independently of alien causes determining it" (GMS 446; *cf.* KPVV 33 and KRV A543/B 562).[1] To say that the free will operates "independently of alien causes" is not, however, to say that it does not operate according to any cause. On the contrary, since the will is a cause, and Kant thinks that all causes must act in accordance with laws, even the free will must act in accordance with some laws. So, in order for a will to be free, it must operate according to laws that it imposes on itself. This leads to Kant's positive characterization of freedom as "the will's property of being a law to itself" (G IV447; *cf.* KPV V33 and MS VI214). Equivalently, a free will is an autonomous will.

Now, in GMS II, Kant had argued that for a will to act autonomously is for it to act in accordance with the categorical imperative,

I thank Des Hogan, Tim Stoll and Dimitris Tsementzsis for this discussion.

1 All translations are Gregor (1996).

the moral law. Thus, Kant famously remarks: "a free will and a will under moral laws is one and the same" (*ibd.*) and therefore to show that the moral law applies to all rational beings, it will suffice to show that those beings are free: "If, therefore, freedom of the will is presupposed, morality together with its principle follows from it by mere analysis of its concept" (*ibd.*).

The question of whether freedom of the will can be presupposed is taken up in the next part of section three, which is titled "Freedom Must be Presupposed as a Property of the Will of All Rational Beings." After setting up the task as that of giving an a priori proof of freedom of the will (GMS IV 447-448), Kant claims that if a rational being must view itself as free, or "must act under the idea of freedom", then that being is free in a practical respect. By referring to the kind of freedom that he wants to show that all rational agents have as "freedom in a practical respect", Kant is distinguishing his project from that of showing that every rational agent is in fact free, which Kant refers to as "freedom in a theoretical respect" (*cf.* Kant's footnote to IV 448). To show that all rational agents have freedom in a theoretical respect would be to show that all rational agents are free. To show that all rational agents are free in a practical respect, by contrast, would be merely to show that all rational agents must regard themselves as free. The reason why it is sufficient for Kant's purposes to prove only that all rational beings are free in a practical respect is that being free in a practical respect means being committed to viewing the moral law as applying to oneself. For, as we said before, to be free is just to act in accordance with the moral law.

Thus, the crucial part of the argument is the next step, in which Kant argues that all rational beings are free in a practical respect. It is in this portion of the argument that the key passage that I want to deal with occurs. Kant begins by asserting that for "[every rational being] we think of a reason that is practical, that is, has causality with respect to its objects" (GIV 448). The idea that the will is a "causality with respect to objects" is meant to suggest that to have practical reason is to have a will that is capable of generating reasons for action from itself.[2] Here Kant is making the same distinction we saw at the beginning of the section with respect to freedom. What Kant wants to say is that in order to view oneself as having a practical will, one must regard one's will itself as generating motivation for acting. If one only views oneself as acting on the basis of impulses, then one cannot also view oneself as having a practical will.

2 The way Kant phrases it is somewhat difficult because he seems to imply that every will is a practical will. Regardless of whether Kant really thinks this, I will take him here to be talking about a specifically practical will.

Now, the passage that I want to focus on throughout the rest of this essay occurs immediately after the sentence just quoted. I quote it in full, labeling the two sentences for clarity in my later discussion:

> [1] Now, one cannot possibly think of a reason that would consciously receive direction from any other quarter with respect to its judgements, since the subject would then attribute the determination of his judgment not to his reason but to an impulse. [2a] Reason must regard itself as the author of its principles independently of alien influences; [2b] consequently, as practical reason or as the will of a rational being it must be regarded of itself as free, that is, the will of such a being cannot be a will of his own except under the idea of freedom, and such a will must in a practical respect thus be attributed to every rational being. (GIV 448)[3,4]

Now, Sentence 2 states the conclusion of the argument, and specifically, the part of the conclusion that is of most interest for our purposes, is 2a, so let us begin our analysis of this passage by considering that. What does Kant mean when he says, "Reason must regard itself as the author of its principles independently of alien influences?" There are three questions here: (1) What is Kant referring to when he talks about "Reason"?; (2) What is it to say that "reason must regard itself as ...?"; and (3) What is it for reason to be "the author of its principles independently of alien influences?"

Regarding (1), the context of the argument is, of course, a discussion specifically of practical reason. So one might think that Kant is making a point here only about the practical domain. But I believe it is fair to think that Kant is talking about both practical and theoretical reason based on three considerations. First, Kant states the argument in terms of reason as such, so, taken at face value, it seems obvious that we ought to take him as speaking not only about practical reason, but about theoretical reason as well. Second, when one considers the argument, it seems clear that if it is sound, it will apply just as much to theoretical reason as to practical reason. So, even if Kant had framed the argument specifically in terms of practical reason, it would still be reasonable to extend the argument to apply to the theoretical case

3 *„Nun kann man sich unmöglich eine Vernunft denken, die mit ihrem eigenen Bewußtseyn in Ansehung ihrer Urtheile anderwärts her eine Lenkung empfienge, denn alsdenn würde das Subject nicht seiner Vernunft, sondern einem Antriebe, die Bestimmung der Urtheilskraft zuschreiben. Sie muß sich selbst als Urheberin ihrer Principien ansehen, unabhängig von fremden Einflüssen, folglich muß sie als practische Vernunft, oder als Wille eines vernünftigen Wesens, von ihr selbst als frey angesehen werden; d.i. der Wille desselben kann nur unter der Idee der Freyheit ein eigener Wille seyn, und muß also in practischer Absicht allen vernünftigen Wesen beygelegt werden.*"

4 All German quotations are from the *Akademie Ausgabe*.

as well.[5] And third, Kant does at least in one point make this same argument in specific connection with theoretical reason, namely in the review of Schulz, which will be discussed below.

So much, then, for what the intended referent of "reason" here is. Now, regarding (2) there seems to be two questions that need to be answered: a) what, in general, is meant when Kant says that *x* must regard itself as *y*? And b), what does it mean for reason to have to regard itself as whatever? I shall have relatively little to say in response to a) other than to point out the difference between the claim that reason must hold that it is free and the claim that reason must regard itself as free. The difference between the two might elude the casual reader, insofar as locutions like the following seem to mean roughly the same thing: "I regard her as the best athlete on campus" and "I think that she is the best athlete on campus." But it is clear that in this context the claim that reason must regard itself as *F* is not equivalent to the claim that reason must hold that it is *F*. This is seen from the fact that Kant is very careful to make clear that he does not intend to give a theoretical proof, but rather a practical proof of freedom of the will. The difference between these two kinds of proof is precisely the difference between showing that some claim is true and showing that we must act or think as if some claim is true, where this leaves open the theoretical question of whether the claim is true or not.

What it is to regard *x* as *y*, in Kant's sense, then, seems intuitive enough. But why does Kant say that reason must regard itself as free instead of that we must regard ourselves as free? One way to gloss this claim might be to say that insofar as we are rational, we must regard ourselves as free. But I think this is unsatisfactory since it seems to imply that we are rationally obligated to think that we are free. But if Kant has shown that we are rationally obligated to believe that we are free, then he seems to have given a theoretical proof of the proposition rather than a practical proof. A better way to gloss Kant's conclusion, then, would be to say that insofar as we take ourselves to be rational, we must take ourselves to be free. This gloss would explain why Kant frames his conclusion in the immediately surrounding context as that freedom is a property of all rational beings. The thought is that to be rational is to be free in a certain sort of way so if one regards oneself as rational, one must regard oneself as free.

Having said this much, we are able to provide a rough sketch of what the argument in this passage is supposed to be.

1. What it is for a will to be free is for it to act autonomously (the positive definition of freedom), or equivalently, for it to act independently of alien influences (the negative definition of freedom).

5 This point is also made by Mieth and Rosenthal (*cf.*254).

2. To take oneself to be rational is to take oneself as able to act or believe independently of alien influences.
3. Therefore, to take oneself to be rational is to take oneself to act or believe independently of alien influences.

The main interpretative question that remains for us is, then, to explain why does Kant endorse 2. And to answer this, we must answer the question of what Kant means when he refers to reason being independent of alien influences.

By "alien influences" (*fremden Einflüssen*), Kant seems to mean any sort of impulse that has its origin in the natural, or in Kant's words, "phenomenal" world. Thus, Kant is providing here a kind of argument against determinism: we cannot rationally believe that determinism extends so far as to include even our beliefs and rational assessments, because if we do, then we abolish the possibility of regarding any of our beliefs as rational. To see why Kant thinks this, let us turn to Sentence 1.

Why Does Theoretical Rationality Imply Freedom?

The thought in Sentence 1 is, I think, best illustrated by an example. If I form the belief that quiche is an unhealthy food and realize upon reflection that the only reason for my belief is a dislike for its taste, then I cannot regard the belief that quiche is unhealthy as rational. In this case it was not my rational faculty, or as Kant simply puts it, "reason" that was active in forming the belief at all. It is not the case that one way of using one's reason is by being guided by impulses in this way; rather, this is a way of not employing one's reason. This is why Kant says that "one cannot possibly think of a reason" (IV 448) that would be affected in this kind of way.

So, if we want to understand Kant's argument, we ought to begin by analyzing cases such as these. Let us focus on the following case. Suppose Ned has some belief *p* and comes to believe that the reason that he believes *p* is because of, say, the presence of a certain chemical in the drinking water. In this case, the rational basis that Ned has for believing the proposition in question is completely undercut. Upon finding out the unsavory origin of his belief, he is rationally obligated to decrease his credence in the proposition. So, one way to make sense of Kant would be to find the principle that underlies the fact that Ned cannot rationally take himself to be rationally justified in believing *p* and explain why one might think that this principle applies to all of one's beliefs if determinism is true. So, what might this principle be? One, I think, clearly unsatisfactory answer might be something like the following:

(P1) If *s* discovers that he believes *p* purely based on material causes, then *s* cannot take himself to be rationally justified in holding that belief.

This principle, if true, would of course get Kant to his desired conclusion. But it is problematic for several reasons. First, the fact that the cause of the belief in Ned's case is the outcome of material causes is irrelevant to whether Ned is justified in his belief. To see this, we can just modify the case so that instead of a chemical in the drinking water, it is an immaterial demon that is responsible for implanting the belief in Ned's mind. And according to that same case, we can see that something like the following principle cannot do the necessary explanatory work either:

(P2) If *s* discovers that she believes *p* purely based on deterministic causes, then *s* cannot take herself to be rationally justified in holding that belief.

This principle cannot be correct either because we can just suppose that the demon responsible for implanting the belief had free will and implanted the belief in one's mind. But, even more importantly, there seem to be plenty of obvious cases of one's holding a belief based on purely material or deterministic causes, where this fact in no way undermines the rational credibility of the belief. For example, I have the belief that there is a book on my desk because I see it. And seeing the book seems to be a purely material and deterministic process. So the real reason why Ned is not justified in holding his belief cannot be sufficiently explained by the mere fact that his belief had a material origin or that it was the outcome of deterministic processes.

A more promising approach would stress the fact that in Ned's case, if *p* does turn out to be true, this would be an example of mere epistemic luck. Although he has no reason to think that *p* is true, he also has no reason to think that *p* is false. Thus, one might think that the reason why Ned would not be justified in believing *p* has to do with the fact that his belief was caused by factors that do not bear on the truth of *p*. So, we might use the following principle:

(P3) If *s* believes *p* as a result of factors that do not bear on the truth of *p*, then *s* is not justified in believing *p*.

In order for Kant to get to his desired conclusion, then, he would need, in addition to (P3), something like:

(P4) If *s* believes that *p* as a result of deterministic processes, then *s* believes *p* as a result of factors that do not bear on the truth of *p*.

The interpretive approach of attributing this principle or something like it to Kant is favored by Jordan who writes the following about Kant's argument:

> Suppose we are asked to accept the proposition that all our rational assessments have sufficient –not just necessary– causal conditions. In order to show that we ought to believe this, someone would need to produce evidence which is seen to conform to criteria of reasonable trustworthiness and which is recognized to confer [. . .] certainty or sufficient probability upon it. But if the proposition is true, this could never happen, for it implies that whether anyone believes it and what he considers trustworthy evidence and acceptable principles of inference are determined altogether by conditions that have no assured congruence with the proposition's own merits or with criteria of sound argumentation whose validity consists in more than that we accept them. (53-54)

The principle that Jordan attributes to Kant, then, is that if determinism is true, our rational assessments of a proposition "have no assured congruence with the proposition's own merits", and this is roughly equivalent to (P4).

What, then, are we to make of (P4)? This principle does not have much intuitive appeal. Consider again the case of beliefs formed on the basis of sense-perception. The reason I believe, say, that there's a book on the table is because the book has impacted my visual system in some way, and had the book not been there, it would not have impacted my visual system in that way. So, it seems clear that in that kind of case, the reason why I believe that there's a book on the table is both the outcome of certain deterministic processes, and my believing it is correlated with the truth of the proposition. And it's not clear why other cases of rational assessment would be any different. For example, a calculator functions in a deterministic way, but there's no reason to think that a calculator's results "have no assured congruence with the [output's] own merits." Why couldn't we think of human minds in roughly the same way? That is, why couldn't we think that the mind is just like a very complex calculator or computer, and that humans often get things right (*i.e.* by producing accurate judgments) in the same way in which calculators and computers often get things right (*i.e.* by producing accurate outputs)?

Now, one possible reason why one might think that there's a difference between the two cases might have to do with the nature of the connection between the fact's being true and the agent believing the proposition that is required for that belief to be justified. Consider again the Ned case. Suppose that *p* is the claim that the drinking water

contains chemicals that can affect one's beliefs. And suppose, as in the original Ned case, that Ned believes this purely because the chemicals in the drinking water have affected his brain in such a way that he comes to believe the proposition, but without knowing why. In this case, upon discovering the true origin of the belief, Ned is rationally obligated to view himself as having been rationally unjustified in holding it. The fact that he believes *p* is a result, in part, of the truth of *p*.

So, a better account of why Ned cannot be rationally justified in the original case might appeal to the thought that if one is rationally justified in believing something one must have the right sort of access to the reasons why one believes the proposition, such as, for example, to be able to explain why one holds the view, upon being prompted. Thus, we might state the principle as follows:

(P5) If *s* is unaware of why she believes some claim *p*, then *s* cannot take herself to be rationally justified in believing *p*.

In order for Kant to get to his desired conclusion, he would need something like the following principle:

(P6) If determinism is true, then no one believes as the result of grasping reasons for believing.

So, this could help explain why one would think that if determinism is true, we cannot be rationally justified in our beliefs, because you might think that being justified in a belief requires a sort of awareness of the reasons why one believes what one does. And one might think that insofar as determinism implies that our beliefs are caused by material processes, which we don't have direct mental access to, determinism implies that we are not aware of the reasons for our beliefs in the right sort of way. How exactly this argument is supposed to work, though, will require closer attention. So let us begin by examining (P6) in somewhat more detail.

This principle does seem to have some intuitive appeal, and I think there is evidence to believe that this is something like what Kant had in mind. Let me begin by discussing the evidence for attributing this view to Kant, and then I'll comment on the philosophical tenability of this position.

Kant advances a very similar argument to the one under consideration from GMS III in a review of a book by Johann Heinrich Schulz (1783), just two years before Kant completed the manuscript of the GMS. After a summary of Schulz's book (*cf.* RS VIII 10-12) and a few sympathetic remarks *(cf. id.* 12-13), Kant writes: "However, no impartial reader [...] will fail to note that the general fatalism which is the most prominent principle in this work and the most powerful one, affecting all morality,

turns all human conduct into a mere puppet show" (*id*. 13). Kant goes on the next page to raise an argument against fatalism:

> Although he [Schulz] would not himself admit it, he has assumed in the depths of his soul that understanding is able to determine his judgment in accordance with objective grounds that are always valid and is not subject to the mechanism of merely subjectively determining causes, which could subsequently change; hence he always admits freedom to think without which there is no reason. (RS VIII 14)[6]

The thought here is that if determinism is true, we must regard our beliefs or rational assessments as "subject to the mechanism of merely subjectively determining causes", whereas in order to regard ourselves as rationally justified in holding some belief, we must regard that belief as a result of "the judgment in accordance with objective grounds that are always valid." By "subjectively determining causes," I take it, Kant is referring to the mechanisms through which one came to hold some belief. These are subjective in the sense that they apply by virtue of facts about the situation that could fail to obtain. But why couldn't someone think that a belief is both the result of subjectively determining causes in this sense and also in accordance with objective grounds? Wood offers the following analysis of Kant's argument in this passage:

> Fatalism [*i.e*. the view that "our actions are necessitated by the mechanism of nature and that this precludes practical freedom"] would have no philosophical interest if fatalists could not also represent themselves as denying freedom for good reasons. [...] But that means they must already presuppose both in themselves and in others the capacity to act according to rational norms in settling theoretical questions. That capacity, however –the capacity to act according to norms of reason– presupposes freedom. It follows that fatalists must presuppose the contradictory of what they are trying to prove even in undertaking to prove it. (2007 131-132)

So, on Wood's view, the reason why we cannot think of a belief as both the outcome of subjectively determining causes and as based on objective grounds is that in order to regard a belief as based on objective grounds, one must regard oneself as free. This analysis puts us in a position to formulate an initial argument for (P6):

1. If *s* believes *p* on the basis of reasons, *s* is free.
2. If determinism is true, then no one is free.

6 *„Er hat aber im Grunde seiner Seele, obgleich er es sich selbst nicht gestehen wollte, voraus gesetzt: daß der Verstand nach objectiven Gründen, die jederzeit gültig sind, sein Urtheil zu bestimmen das Vermögen habe und nicht unter dem Mechanism der blos subjectiv bestimmenden Ursachen, die sich in der Folge ändern können, stehe; mithin nahm er immer Freiheit zu denken an, ohne welche es keine Vernunft giebt."*

3. Therefore, if determinism is true, one never believes on the basis of reasons.
4. Therefore, if determinism is true, then no one believes as a result of grasping reasons for believing P6.

Premise 2 follows from Kant's incompatibilism, so the only premise that requires attention for our purposes is 1. And, in order to understand why Kant endorses 1, we need to understand what it means, according to Kant, to believe on the basis of reasons.[7] One intuitive way of glossing what it is to believe on the basis of reasons is to say that the reasons for believing the proposition caused the belief in some way (*i.e.* they mediately caused the belief through the agent's grasping or recognizing them). However, although Kant thinks that reasons may explain the fact that an agent believes something, he does not believe that reasons can cause a belief. This is because, according to Kant, for *x* to cause *y*, there must be some law according to which whenever *x* is present, *y* follows of necessity.[8] By contrast, it cannot be the case that the fact that there are good reasons for acting or believing, or even that the agent recognizes that there are good reasons for acting or believing is itself sufficient for an agent acting or believing in that way. The fact that I have good reasons to act or believe in a certain way will not itself bring it about that I do believe or act in a certain way. For example, if I'm a weak-willed agent I might act contrary to what I know to be the course of action that I have most reason to do. Thus, as Leibniz says, reasons merely "incline without necessitating" (128; *cf.* Locke book II, chapter XXI, §§47, 51-52, 71, 263-266, 282-284).

Now the epistemic case is a bit more complicated in that it does seem that recognizing reasons for belief can be sufficient to generate a belief. It doesn't seem inconceivable, though, that one could genuinely recognize (decisive) reasons for believing and yet still not believe, for example, the psychological phenomenon referred to as denial. It seems, then, reasonable to think that reason, in the epistemic case, is the same as in the practical case.

So, given this account, we can formulate an argument for (1) along the following lines:

1. If *s* believes *p* on the basis of good reasons, then reasons for believing *p* explain the belief in some way.

7 On this account of believing for reasons, I follow Wood (*cf.* 127-128).

8 See, for example, his characterization of causation in the first Critique: "Every cause pre-supposes a rule according to which certain appearances follow as effects, and every rule requires a uniformity in its effects" (A 549/B 577).

2. If reasons explain the belief, then it cannot be the case that material conditions are sufficient to explain the belief.
3. If material conditions are not sufficient for the formation of the belief, then one could have believed otherwise.
4. Therefore, if one believes *p* on the basis of reasons, then one could have believed otherwise.
5. If one could have believed otherwise, then one is free.
6. Therefore, if *s* believes *p* on the basis of reasons, then *s* is free.

Premise 1 falls out of the analysis of what it is, according to Kant, to believe on the basis of reasons. Premise 3, likewise, seems straightforward: if it is not necessitated that I do φ, then I could have failed to do φ. And 5 follows from the plausible principle that one can act otherwise if and only if one is free to do so.

Premise 2 is a little more difficult, but I think it has a good basis in Kant as well. Consider, for example, the following passage in the first *Critique*:

> It is easy to see that if all causality in the world of sense were mere nature, then every occurrence would be determined in time by another in accord with necessary laws, and hence —since appearances, insofar as they determine the power of choice, would have to render every action necessary as their natural consequence— the abolition of transcendental freedom would also simultaneously eliminate all practical freedom. For the latter presupposes that although something has not happened, it nevertheless ought to have happened, and its cause in appearance was thus not so determining that there is not a causality in our power of choice such that independently of those natural causes and even opposed to their power and influence it might [...] begin a series of occurrences entirely from itself. (KrV A534/B562)[9]

The thought here, I take it, is that if an account of action in terms of that action's deterministic causes is complete, there will be no room left to explain normative facts like whether the action ought to have been done. If considerations of what reasons the agent had for acting have a place in the explanation of the action, then the action must not be determined in the strong sense (the causes were "so determining") that deterministic explanation is fully sufficient to account for the action.

So, the underlying principle behind 2 is, I think, intuitive enough. We might state it as:

(P7) If *p* is a sufficient condition for *r*, and *q* is distinct from *p*, then *q* cannot be a necessary condition for *r*.

9 Guyer and Wood translation (1998).

An illustration of why one might think that this is true might be the following. If my paying the relevant amount of money to buy a car from a dealership is a sufficient condition for my obtaining the car, then it cannot also be the case that making a good impression on the salesperson is a necessary condition for me to obtain the car. If I pay for it, I'll get the car regardless of whether I make a good impression. Likewise, it is not *prima facie* implausible to think that if neurophysiological states are a sufficient condition for me to believe *p*, then it cannot also be the case that the presence of good reasons for believing *p* is also a necessary condition for my believing *p*. And, if you add the principle used above that if a belief is not caused by the recognition of reasons for believing, then the belief is not justified, then one gets the conclusion that if neurophysiological states are sufficient condition for believing, then no beliefs are justified.

Is Kant's Argument Sound?

Nonetheless, upon further reflection, I think this principle can be decisively seen to be false. Consider the following example. If the car in the above case costs $20,000 dollars, then having $25,000 dollars at my disposal is a sufficient condition for me to be able to buy the car. But it is still true that having $20,000 dollars at my disposal is a necessary condition for me to be able to purchase the car, even though these are distinct properties. The logic here is simple. If *p* implies *q*, and *p* is a sufficient condition for *r*, then even if *p* and *q* are distinct properties, *q* is also a necessary condition for *r*. Goldman makes the same point when he introduces the notion of simultaneous nomic equivalents, which he characterizes as follows: If there is a law such that

> for any object of kind H at time *t*, the object has property Ø at t if and only if it has the property Ψ at *t*[,] then if a particular object a has properties Φ and Ψ at a particular time *t1*, I shall say that a's having Φ at *t1*, is a "simultaneous nomic equivalent" of a's having Ψ at *t1*. (473)

So, if a's having Ψ and a's having Φ are simultaneous nomic equivalents, then although a's having Ψ is a sufficient condition for a's having Ø, a's having Φ is a necessary condition as well.

How does all of this relate to the issue of whether one can think of a belief as both based on good reasons and the product of deterministic processes? Let Ψ be the property of having good reasons to believe *p*, and let Φ be the property of being in a certain brain state that is such as to cause (necessitate) one to believe *p*. Let us also grant with Leibniz and Kant that reasons act as a cause of beliefs or actions in a different way than, on the deterministic picture, brain states or whatever do.

This suffices to show that a's having Ψ and a's having Φ are not the same. But it does not at all suffice to show that Ψ and Φ are not, in Goldman's terms, simultaneous nomic equivalents. And if they are simultaneous nomic equivalents, it is open to the determinist to say that although mental states are the result of sufficient material conditions, the recognition of good reasons is at least sometimes causally necessary to produce a belief.

So, suppose determinism is true, and for every belief there is some set of material conditions C that is responsible for my holding that belief. It does not follow that I do not hold that belief on the basis of good reasons because it may be the case that C entails C*, that the agent recognizes good reasons for holding that belief. In that case C can be sufficient for the belief, and C might be non-identical to C*, but nonetheless C and C* are simultaneous nomic equivalents, and thus C* is necessary for the belief.

Now, Hasker has tried to answer Goldman's objection by appealing to the fact that, in order for one to be rationally justified in holding some belief, that belief must not only be held on the basis of reasons, it must be held on the basis of good reasons. He writes:

> On Goldman's view, my accepting such and such as a reason, which may in fact happen to be a good one, is a necessary condition of my reaching a particular conclusion. But the reason's being a good one is not a necessary condition. [...] But for me to be justified in holding a belief surely means that I hold it because I have good reasons, and would not hold it otherwise. (179)

I see no reason, however, why Goldman's account could not accommodate the thought that sometimes people believe things because they recognize good reasons for doing so. There's nothing to prevent Goldman from saying that reasons cause belief and sometimes good epistemic agents are sensitive to whether the reason in question is a good one. A good epistemic agent (qua good epistemic agent) will only believe on the basis of good reasons, and, therefore, in cases where someone is behaving as a good epistemic agent, the goodness of the reasons for belief are relevant.

A Remedy for Kant's Argument

So, I think that if this analysis of Kant's argument is right, the argument is a bad one. I want now to briefly explore how the argument might be changed so as to avoid the difficulty raised by the possibility of simultaneous nomic equivalents. One way to do this is to modify the view that Kant's argument is meant to rule out. In the foregoing, I have taken it for granted that the view that Kant is directed against is

fatalism or determinism.[10] I think, though, that if we take it not to be an argument that regarding ourselves as rational is incompatible with determinism simpliciter but rather, determinism and a certain view about the explanation of mental events, the argument turns out to be much more plausible.

Before I discuss that principle, though, I should note that there's an exegetical issue here as to whether the assumption that Kant is directing his argument simply against determinism is true. I have not defended this assumption because I think it is the most natural way to read the texts with which I have been dealing, and trying to settle the exegetical issue would be beyond the scope of the essay. But I do not think that it is obvious that there is not a more complex dialectic going on here. So, depending on what the claim that one takes Kant to be arguing against is, Kant might have been arguing against something like the position I am about to sketch.

The principle that I have in mind is what Kim calls the "Principle of Explanatory Exclusion" and formulates as follows: "No event can be given more than one complete and independent explanation" (79). So, let E be a mental event –the formation of a belief, for example– and let C be the material conditions that cause the belief and C* the rational conditions that cause the belief. Call C the material cause and C* the rationalizing cause. If there are simultaneous nomic equivalents then both C and C* could be complete and in-dependent explanations of the fact that the agent has formed the belief. A material explanation would be complete, if it were not necessary to refer to any other material event or fact in order to explain the event. Likewise, a rationalizing explanation would be complete if there was no consideration or weighing of a consideration that is necessary to explain why she formed the belief. These two kinds of explanations could be independent in that it is not necessary to refer to rationalizing facts in order to explain the material facts and vice versa.

Suppose, then, that one endorses the Principle of Explanatory Exclusion and thinks that mental events are explicable by reference to material conditions. Either that person must identify responding to reasons as part of the material explanation of the belief, or they must say that reasons do not figure in the complete explanation of the belief.

10 The terms "fatalism" and "determinism" are today usually used to denote different views. Determinism, roughly, denotes the view that all events are necessitated by the conjunction of the laws of nature and the initial state of the universe, or simply laws of nature and antecedent states of the world, whereas fatalism is used to denote the view that we cannot act other than we do act. But I take it, on the basis of Wood's characterization of fatalism above, that the term had a broader usage at Kant's time, so I follow that usage and take determinism and fatalism to be roughly equivalent.

We have already seen a Kantian argument against thinking that responding to reasons can be part of the material explanation. The thought there was that all material causes necessitate their effects, but reasons never function this way, so responding to reasons cannot be part of the material cause. Therefore, the proponent of the Principle of Explanatory Exclusion who believes determinism is true is left with the conclusion that reasons do not figure into the complete explanation of why someone holds a belief. Now if we add the principle that in order for a belief to be justified it must be caused in some way by the recognition of reasons for believing, we have an argument for the claim that if determinism and the Principle of Explanatory Exclusion are true, then no beliefs are rationally justified. We can formulate the argument as follows:

1. Reasons cannot explain beliefs in the same way that material conditions do.
2. Therefore, reference to reasons are not necessary in a complete material explanation of a belief.
3. Therefore, if determinism is true and the Principle of Explanatory Exclusion is true, reasons do not explain beliefs.
4. If *s* believes *p* on the basis of reasons, then reasons for believing *p* explain the belief in some way.
5. Therefore, if determinism and the Principle of Explanatory Exclusion are true, no one believes on the basis of reasons.

Conclusion

I have argued for a particular interpretation of Kant's argument that regarding oneself as rational implies regarding oneself as free. This argument is, I think, a bad one because it neglects the possibility of what Goodman calls simultaneous nomic equivalents. One way to reformulate Kant's argument so as to avoid this difficulty is by taking him not to be arguing against determinism simpliciter, but against the view that determinism is true and that mental events can only be explained as material phenomena. Taking the argument that way yields an interesting and not implausible argument.

References

Goldman, A. I. "The Compatibility of Mechanism and Purpose." *The Philosophical Review* 78.4 (1969): 468-482.

Hasker, W. "The Transcendental Refutation of Determinism." *The Southern Journal of Philosophy* 11.3 (1973): 175-183.

Jordan, J. N. "Determinism's Dilemma." *The Review of Metaphysics* 23.1 (1969): 48-66.

Kant, I. *Gesammelte Schriften.* [AK.]. Hrsg.: Bd. 1-22. Preussische Akademie der Wissenschaften, Bd. 23. Deutsche Akademie der Wissenschaften zu Berlin, ab Bd. 24 Akademie der Wissenschaften zu Göttingen. Berlin: de Gruyer, 1900.

Kant, I. "Groundwork for the Metaphysics of Morals." *Practical Philosophy (The Cambridge Edition of the Works of Immanuel Kant).* Ed. and trans. Mary J. Gregor. Cambridge: Cambridge University Press, 1996. 37-108.

Kant, I. "Review of Schulz's Attempt at Introduction to a Doctrine of Morals for all Human Beings Regardless of Different Religions Part I." *Practical Philosophy (The Cambridge Edition of the Works of Immanuel Kant).* Ed. and trans. Mary J. Gregor. Cambridge: Cambridge University Press, 1996. 3-10.

Kant, I. *Critique of Pure Reason (The Cambridge Edition of the Works of Immanuel Kant).* Ed. and trans. Paul Guyer and Allen Wood. Cambridge: Cambridge University Press, 1998.

Kim, J. "Mechanism, Purpose and Explanatory Exclusion." *Philosophical Perspectives* 3 (1989): 77-108.

Leibniz, G. W. "Discourse on Metaphysics." *Philosophical Essays.* Eds. Roger Ariew and Daniel Garber. Cambridge: Hackett Publishing Company, 1989. 35-68.

Locke, *J. An Essay Concerning Human Understanding (The Clarendon Edition of the Works of John Locke.* Ed. Peter H. Nidditch. Oxford: Oxford University Press, 1979.

Mieth, C., and Rosenthal, J. "'Freedom must be Presupposed as a Property of the Will of All Rational Beings' (GMS III, 2)." *Groundwork for the Metaphysics of Morals.* Eds. Cristoph Horn and Deiter Schoenecker. Berlin; New York: de Gruyter, 2006. 247-284. https://doi.org/10.1515/9783110204551.4.247

Schulz, J. H. *Attempt at Introduction to a Doctrine of Morals for all Human Beings regardless of different Religions Part I.* Berlin: Christian Ludewig Stahlbaum, 1783.

Wood, A. *Kantian Ethics*. Cambridge: Cambridge University Press, 2007. https://doi.org/10.1017/cbo9780511809651