The Self-Consciousness of Instrumental Reason as the Key to Section 2 of *Groundwork III*

Despite its fame and importance, *Groundwork III* does not have many defenders. Its influential interpreters have deemed it a failure, and even the most sympathetic commentators have tended to think that important parts of it contain pre-critical holdovers that are best read as anticipations of Kant’s mature position rather than expressions of it.¹ There are, of course, good reasons for the state of interpretative

¹ See, for example, the following: Allison, Henry: *Kant’s Theory of Freedom*. Cambridge and New York. 1990; Henrich, Dieter: “The Deduction of the Moral Law: The Reasons for the Obscurity of the Final Section of Kant's *Groundwork of the Metaphysics of Morals*”. In: *Kant's Groundwork of the Metaphysics of Morals: Critical Essays*, Paul Guyer, ed. Lanham, Md. 1998, 303-343; Paton, H.J.: *The Categorical Imperative: A Study in Kant’s Moral Philosophy*. Philadelphia, PA. 1971, [c. 1947]; Ameriks, Karl: “Kant’s Deduction of Freedom and Morality”; “Kant’s *Groundwork III* Argument Reconsidered”; “Pure Reason of Itself Alone Suffices to Determine the Will”. In: *Interpreting Kant’s Critiques*. Oxford and New York. 2003, 161-193; 226-249; 249-263. Allison, Henrich, Paton, and Ameriks all consider *Groundwork III* to be a failure. Henrich and Paton believe that all the material for Kant’s mature position are already contained in the *Groundwork*. They argue that Kant himself, while writing the *Groundwork*, came to see that his ambition of deriving freedom from non-moral premises is incoherent. Ameriks differs from Henrich and Paton in thinking that Kant’s ambition of deriving freedom from non-moral premises is not only coherent but also necessary for the completion of Kant’s project. He too, however, thinks that the argument Kant actually mounts is both a failure and a reversion to pre-critical beliefs about freedom. Thus, *Groundwork III* is “a slightly more self-conscious continuation of some relatively crude beliefs about freedom that Kant… simply had not gotten around to submitting to a thorough critique” (ibid. 181). Allison agrees with Ameriks that some argument is required, and agrees that *Groundwork III* fails to provide it. He is more optimistic, however, about Kant’s ability to provide the necessary argument in the *Critique of Practical Reason*. Recently, it has become popular to achieve a certain amount of optimism about *Groundwork III* by arguing that Kant’s central ambition was never to prove autonomy from rationality. On this view, showing that rational agents are autonomous is trivial, and the real heavy lifting only comes about when Kant tries to show that even imperfectly rational agents like us are subject to the moral law. [See, for example, McCarthy, Michael: “Kant’s Rejection of the Argument of *Groundwork III*”. In: *Kant-Studien* 73. 1982, 169-90; and “The Objection of Circularity in *Groundwork III*”. In: *Kant-Studien* 76. 1985, 28-42; Schönecker, Dieter: *Kant: Grundlegung III: Die Deduktion des Kategorischen Imperativs*. Freiburg. 1999, and “How is a Categorical Imperative Possible?” In: *Groundwork for the Metaphysics of Morals*, Christopher Horn and Dieter Schönecker eds. Berlin. 2006, 301-325] This view will be criticized in what
suspicion that surrounds *Groundwork* III. Consider, for example, the second section of *Groundwork* III, entitled “Freedom Must Be Presupposed as a Property of the Will of All Rational Beings,” which some commentators think contains the central argument of *Groundwork* III. What the structure of Kant’s overall argument appears to demand at this point is an argument connecting rational agency and freedom. What we appear to get instead is an argument connecting rationality and the (theoretical) capacity to draw inferences based on reasons. Moreover, Kant abruptly shifts from a discussion about how rational agents are, to a discussion about how rational agents must regard themselves, and seems to think these two types of discussions amount to the same thing when considered “in a practical respect [in praktischer Rücksicht]” (GMS: 448). Finally, to make matters worse the conclusion of section 2 seems to be importantly qualified if not openly disavowed in the paragraphs immediately following it.

Since Kant himself compresses the argument of section 2 to 7 sentences, any interpretation must involve some speculative filling in of gaps. The interpretation offered here will be no exception to this limitation. I will make exegetical progress, despite this follows, but for now it should be noted that, even Schönecker, its most elaborate proponent, is forced to concede that on this interpretation all imperatives are synthetic, thus abandoning an important position from *Groundwork* II. [See Schönecker, *Kant: Grundlegung* III, 90ff.] The one exception to this interpretative evaluation is Mieth and Rosenthal who in Mieth and Rosenthal: “‘Freedom must be presupposed as a property of the will of all rational beings [...]’” In: *Groundwork for the Metaphysics of Morals*, Christopher Horn and Dieter Schönecker eds. Berlin. 2006, 247-285, argue that although *Groundwork* III is a failure, as Kant conceives it, it can be reconstructed in such a way that the central ambition of deriving freedom from rationality is given a successful argument.

2 “Freiheit muss als Eigenschaft des Willens aller vernünftigen Wesen vorausgesetzt werden.”


limitation, by articulating 3 features of the argument and turning these features into 3 criteria of a successful interpretation of section 2. It will turn out that most of the standard interpretations fail to satisfy one if not all of these criteria. The structure of this paper is determined by this methodology. I will start by articulating the three features of the argument that any successful interpretation of *Groundwork* III must have. Since these features often go against the grain of many contemporary interpretations of Kant, I will spend considerable time explaining and defending them. Finally, in the second section, I offer an argument that satisfies all three of these features in conjunction. Although no attempt will be made to argue for the uniqueness of my speculative interpretation, I believe that it is a sufficiently close extrapolation from the three features of Section 2 that it warrants attribution to Kant himself.

I. Three features of section 2 of *Groundwork* III

A. The structure of section 2 is that it assumes rational agency and seeks to derive the freedom necessary for morality (i.e. autonomy or positive freedom)

In section 1, Kant has argued that being subject to the moral law is equivalent to positive freedom, and that this positive freedom and morality are reciprocal concepts. Any creature that is free in the relevant sense is subject to the moral law and any creature that is subject to the moral law is free. With this identity in place, Kant’s ambition of connecting rationality and morality is finally within reach. All he must do is argue that there is a connection between rationality and freedom, and the connection between morality and reason follows immediately by the identity claim. It is natural to suppose that at some point following section 1, Kant will make such an argument that presupposes rational agency and seeks to derive freedom. Both the section headings as well as the content of section 2 make it clear that some version of this argument from rational agency to freedom takes place in section 2. If we take the section headings to be announcements of the thesis of the sections they head, then the thesis of section 2 is to show that ‘freedom must be presupposed as a property of the will of all rational beings’. Since Kant has long since identified ‘will’ with ‘practical reason’ (GMS: 412, 446), section 2 starts from the assumption of a being with practical reason and seeks to derive freedom as a
property of such a being. Kant confirms this when, in the section, he restates his task as showing that “it [freedom] must be proved as belonging to the activity of all beings whatever that are rational and endowed with a will” (GMS: 448). Section 2, therefore, boldly announces as its theme precisely the argument that we might expect Kant to adopt in order to fulfill his long-standing ambition of deriving morality from reason.

But, if it is clear that ‘freedom’ and ‘rational being with a will’ are the names of the ending point and the starting point of Kant’s argument, it is less clear what Kant means by these terms. For Kant, a rational being with a will (GMS: 412) is that sort of being that has the power of rational causality (GMS: 446). If we understand the power of rational causality to be the power to effect things in the world based upon our reasons, then a living being with a will is just a being for whom reasons are causes. For such a being, the explanations for her actions are simply the reasons that she used in deciding what to do. If Rosemary decides to go to the zoo because she has been working too hard, then it is Rosemary’s power of rational causality that makes it appropriate to explain Rosemary’s action in terms of her reasons. The correct answer to the question ‘Why did Rosemary go to the zoo?’ is not a causal investigation into her brain functioning at the time of decision, but has rather to do with the reasons that Rosemary herself used in making her decision, i.e. ‘She was working too hard’. In such a circumstance, we could also say that Rosemary has exercised practical reason (GMS: 412) or, in contemporary terms, Rosemary has exercised her rational agency. For Kant, then, having a will, being an agent, and having practical reason are all just ways of talking about rational causality. According to the first part of the first criterion, such a being is the starting point of the argument in section 2.

Although Kant’s understanding of a rational being with a will is relatively univocal, Kant uses the term ‘freedom’ in different ways and we must distinguish between three different senses of freedom that are operative in section 2:

5 “[… ] man muss sie als zur Tätigkeit vernünftiger und mit einem Willen begabter Wesen überhaupt gehörig beweisen.”
6 In footnote 10, I will discuss Henrich’s dissenting opinion that having a will should be identified with what I will shortly call autonomy.
7 Henrich, “The Deduction […]”, 314, similarly distinguishes between three different types of rational beings each of which has a kind of freedom appropriate to it. Indeed, I
A. Logical freedom: It is possible that a rational being has the capacity to reflect on herself and her surroundings, make judgments about the truth or falsity of propositions, but completely lack the capacity to act on these judgments. Such a creature would have a certain sort of freedom because she must be able affirm or deny a predicate based on the fact of the matter rather than being constrained to do so by circumstances. However, this freedom only extends to the theoretical sphere, and such a creature does not have the capacity to effect changes in the world based on her judgments. In other words, it is possible that a theoretically reflective creature is not an agent and consequently has no will.

B. Agentive freedom: If we add agency to the theoretically rational creature imagined in (A) then such a creature would have the capacity to effect changes in the world based on her theoretical judgments. Not only would she have the capacity to form judgments about the world based on her reasons, but she would also have the capacity to form judgments about what she should do based on reasons. She would even be able to act in accordance with these practical judgments. Rosemary demonstrates agentive freedom when she successfully goes to the zoo in order to relax.

C. Autonomy: Now let us add to the rational agent in (B) the capacity not only to adopt the rational means to her ends but also the capacity to set her own ends based solely on rational principles. When Rosemary goes to the zoo in order to relax, her action is proximately determined by reason, but in another sense it is determined by inclination. After all, the end of relaxation toward which going to the zoo is a means is itself given to her, not by reason, but by inclination. If, however, Rosemary were to select not only the means, but also the end based on reasons rather than inclination, if, in other words, the action was fully determined by reason, then, Rosemary would possess not only agentive freedom, but autonomy. Groundwork I and II have further specified what such an autonomous action looks like. Kant explains that heteronomy results whenever the will determines itself by means of “an incentive that the anticipated effect of the action has upon the will: I ought to do something on this account, that I will something else” (GMS: 444). In other words, the will is determined by an alien cause, or heteronomously, if and only if, its reasoning can be captured in an hypothetical imperative. Conversely, an autonomous will is one that is capable of determining itself according to reasoning that is not captured in hypothetical imperatives. If a creature with agentive freedom is one whose actions can be explained by reasons in general, then a creature with autonomy is one whose actions can specifically be explained by non-hypothetical reasons. In contemporary terms, an autonomous agent is one who is capable of acting on non-instrumental reasons.

have adopted the term “logical freedom” from Henrich. Henrich, however, does not give sufficient weight to the second type of freedom—only mentioning it in passing and giving it no role in Kant’s argument.

8 “[…] die Triebfeder, welche die vorausgesehene Wirkung der Handlung auf den Willen hat; ich soll etwas tun, darum, weil ich etwas anderes will […]”
In deciding which of the types of freedoms is the intended conclusion of Kant’s argument, we are aided by the fact that section 2 is immediately preceded by a lengthy discussion of freedom in section 1. It is reasonable to suppose that the freedom, which must be proved in the second section, is the same kind of freedom that is discussed in the first section. Turning to the first section, we find Kant defining “negative freedom” as follows:

Will is a kind of causality of living beings in so far as they are rational, and freedom would be that property of such causality that it can be efficient independently of alien causes determining it. (GMS: 446)\(^9\)

Since rational causality or will is the capacity to effect things in the world based on our reasons, free rational causality would be the capacity of that will to operate “independently of alien causes determining it [unabhängig von fremden sie bestimmenden Ursachen]” (GMS: 446). But what counts as an alien cause determining the will? Certainly, for Kant, empirical desires and inclinations count as causes alien to the will, and the will is determined by an alien cause when inclinations simply overwhelm it. If, against her better judgment, Rosemary finds herself yet again unable to relax at the zoo because anxiety is forcing her to work, Rosemary’s will has clearly been determined by an alien cause. But, as the discussion of “positive freedom” which follows and “flows from [fliesst aus]” this definition, makes clear, it is not only these akratic cases in which the will is determined by alien causes. As Kant goes on to make clear, the will is determined by alien influences whenever it is determined heteronomously (446). It seems clear, therefore, that a will that is not determined by alien influences is none other than an autonomous will.

Indeed, since only autonomy has been shown to be equivalent to the moral law, if Kant is serious about his ambition of using freedom as an intermediate step between rationality and morality, then the sense of freedom that is derived from agency better be the very same sense of freedom that is used to derive morality. In other words, if we understand the second section as carrying forward the themes and arguments of the first section, then Kant’s argument in the second section must show that rational agents,

---

\(^9\) “Der Wille ist ein Art von Kausalität lebender Wesen, sofern sie vernünftig sind, und Freiheit würde diejenige Eigenschaft dieser Kausalität sein, da sie unabhängig von fremden sie bestimmenden Ursachen wirkend sein kann.”
whose reasons can be causes, are necessarily also creatures for whom non-instrumental reasons can be causes.\(^\text{10}\)

While many commentators would agree that it is indeed the ambition of section 2 to prove freedom in the sense of autonomy, most think that Kant fails utterly to achieve this ambition. According to these commentators, when we actually investigate the details of section 2, we find that Kant does not provide a proof that ends in freedom in the strong sense required for morality (i.e. autonomy). Kant does indeed argue from ‘rationality’ to ‘freedom’, but the freedom that is argued for is only logical freedom, or at best agentive freedom, whereas the freedom that is required is autonomy. According to these commentators, Kant’s failure, here is inevitable. While Kant is entitled to assume that we are theoretically reflective creatures, and hence rational beings, it does not follow from this that we are also rational agents. In fact, Kant himself seems to imagine precisely such a theoretically rational creature who lacks a will at GW 400. Moreover, even if we grant to all rational beings agency, it is yet a further step to then argue that all rational agents are autonomous. Kant’s argument therefore, fallaciously assumes 1) that all rational beings have wills and 2) that all rational wills have autonomy.\(^\text{11}\)

\(^{10}\) At GMS: 459, Kant seems to identify a will with a free will, “[…] will, that is, of a faculty distinct from a mere faculty of desire (namely, a faculty of determining itself to action as an intelligence and hence in accordance with laws of reason independently of natural instincts). [eines Willens, d.i. eines von blossom Begehrungsvermögen noch verschiedenen Vermögens (nämich sich zum Handeln als Intelligenz, mithin nach Gesetzen der Vernunft, unabhängig von Naturinstinkten, zu bestimmen) […]” Kant is entitled to make this claim here because he has already accomplished his proof that rational agents with a will are free. In other words, it is only after Section 2 that Kant allows himself to identify wills with free wills. Henrich, “The Deduction […],” 314, use of this passage to show that even in Section 2 Kant has identified the will with an autonomous will is a mistake.

\(^{11}\) Thus, Ameriks says, “Such ‘taking’ [i.e. believing ourselves to be agents] does not even establish real efficacy, and even if it did coincide with such efficacy, it would not demonstrate any claim of negative freedom [i.e. Amerik’s term for what I have been calling autonomy] […]” (“Kant’s Groundwork Argument […],” 244). Allison notes that Kant himself could not have identified rational beings with rational agents (Kant’s Theory of Freedom, 218), and thinks that the purpose of the sections 3 and 4 is precisely to show that rational beings have wills (ibid., 224). Ultimately, however, even if this argument that all rational beings have wills is successful, which Allison doubts, Kant’s argument is still ultimately a failure since rational agency is not sufficient to demonstrate our autonomy. “The main point is simply that given the identification of will and
Turning to Kant’s text, we find that there is ample evidence to support the claim that Kant’s argument does not end with the type of freedom that is required for morality. Section 2 is a single paragraph, containing 7 sentences and 3 distinct parts. In the first part, sentences 1-2, Kant announces the thesis of the section, that rational agents are free, and connects it with the results of section 1. In the second part, sentences 3-4, he claims that it is sufficient to prove the seemingly more modest thesis that rational agents must act under the idea of freedom. In the third part, sentences 5-7, Kant offers his argument in support of the thesis. Since commentators have correctly taken these last few sentences to be the key sentences of the section, I quote them in full:

Now, one cannot possibly think of a reason that would consciously receive direction from any other quarter with respect to its judgments, since the subject would then attribute the determination of his judgment not to his reason but to an impulse. Reason must regard itself as the author of its principles independently of alien influences; 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 be attributed to every rational being. (GMS: 448)

Although, as the course of my argument will show, this passage is quite complex, the passage supports a rather simpler standard interpretation of it. Suppose, for example, that practical reason the claim that rational beings posses a will can mean either merely that reason is practical or that pure reason is practical. The former suffices to show that we are genuine rational agents rather than automata; but the latter is required to establish our autonomy” (Kant’s Theory of Freedom, 228). In “The Deduction […],” 315ff, Henrich insists that Kant does not fallaciously infer autonomy from logical freedom, but he insists that the connection amounts to something less than an argument that all rational beings are autonomous: viz. an analogy. Just as theoretical reason is free to act on its principles, so too practical reason must be free to act on its principles, i.e. autonomy. But perhaps the analogy is simply: as theoretical reason is free to act on its principles, so too practical reason must be free to act on its instrumental principles. In other words, the analogy need not go beyond agentic freedom.

12 “Nun kann man sich unmöglich eine Vernunft denken, die mit ihrem eigenen Bewusstsein in Ansehung ihrer Urteile anderwärtser eine Lenkung empfinge, denn alsdenn würde das Subjekt nicht seiner Vernunft, sondern einem Antriebe die Bestimmung der Urteilskraft zuschreiben. Sie muss sich selbst als Urheberin ihrer Prinzipien ansehen, unabhängig von fremden Einflüssen, folglich muss sie als praktische Vernunft, oder als Wille eines vernünftigen Wesen, von ihr selbst als frei angesehen werden; d.i. der Wille desselben kann nur nter der Idee der Freiheit ein eigener Wille sein und muss also in praktischer Absicht allen vernünftigen Wesen beigelegt werden.”
Rosemary concludes that the zoo is closed based on her investigation of the zoo’s website which posts their hours. In order for Rosemary’s conclusion to be a conclusion of reason, we must believe that reason itself is the determining factor in the formation of Rosemary’s judgment. In other words, we must believe that Rosemary forms the belief that the zoo is closed based on the evidence that she has uncovered on the internet. If, however, Rosemary’s belief is formed on some other grounds (because of her anxiety, say), if, that is, it is determined by some influence alien to reason, then Rosemary is not actually engaged in the process of making a judgment at all. It is only when Rosemary has the freedom to form conclusions based on her reasoning that we can think of Rosemary as reasoning in the first place.

The point here is similar to the point Kant made when he defined rational agency. There, Kant argued that a creature with a will was one whose reasons, rather than something else, were the explanations for her actions. Here, Kant seems to make the same point about reason generally. If we are to attribute reasoning to someone, then the outcome of their reasoning must be explained by their reasons. Here, Kant emphasizes that a reasoner must have the freedom to conclude (either in a theoretical or a practical judgment) as reason directs. To be determined by anything other than reasons is to fail to be a reasoner.

But if the standard interpretation of this passage is correct, then commentators are justified in accusing Kant of failing to provide an argument from rationality to morality. On this interpretation, Kant has done nothing other than point out that rationality must be responsive to reasons. But, by itself this fact offers no guidance at all in how to show that rational beings are agents or that rational agents are autonomous. After all, even someone obeying strictly instrumental reasons must have the capacity to conclude as reason directs. To say that Rosemary has the capacity to act on reasons is perfectly compatible with also saying that all of these reasons can be captured in hypothetical imperatives and are therefore heteronomous.

Against the background of this standard interpretation and its accompanying objection, the full force of the first feature can be appreciated. If I am correct, Kant, in section 2, simply assumes that we are rational agents, and he need not argue for this assumption from a prior assumption of theoretic rationality or anything else. Moreover,
once we see that rational agency, and its accompanying agentive freedom are assumed rather than argued for, the standard interpretation of the last few sentences of section 2 is no longer plausible. But with agentive freedom presupposed, the only stronger type of freedom left to be proved is autonomy. Although such an argument from agentive freedom to autonomy comports well with Kant’s overall ambition, finding such an argument within these three crucial sentences of section 2 is another matter altogether, and it is to this task that that I now turn.

B. The argument is from the point of view of rational agency

Stating the conclusion of his argument Kant remarks that: “reason must regard itself as the author of its principles independently of alien influences; consequently, […] it must be regarded as free [Sie muss sich selbst als Urheberin ihrer Prinzipien ansehen, unabhängig von fremden Einflüssen, folglich muss sie [...] als frei angesehen werden]” (GMS: 448, my italics). It is possible to take the ‘consequently’ to be expressing a relation between reason’s being the ‘author of its principles independently of alien influences’ and ‘freedom’ (in the sense of autonomy). If this is true, then Kant is saying that our reason is autonomous because it is independent of alien influences. Since this reading has Kant simply reiterating the definition from section 1 according to which autonomy is just independence from alien influences, it is not very informative as to Kant’s intentions. But, there is another interpretation that turns out to be richer in its implications. Suppose that the ‘consequently’ expresses a relation between reason regarding itself as autonomous and reason being regarded as autonomous. On this second interpretation, Kant is arguing that we know that rational beings are autonomous because from their point of view, they must regard themselves as autonomous. In the second criterion of a successful interpretation, I turn this more informative understanding of this sentence into a crucial feature of Kant’s argument. An argument that does justice to Kant’s intentions must crucially make use of the fact that it is only because reason must view itself as autonomous that we are entitled to conclude that it is autonomous. This does nothing of course to explain why reason must regard itself as autonomous. It merely insists that Kant’s argument for autonomy will rely upon the perspective of reason in order to carry forward its claim.
The passage cited above is by no means the only indication of the importance of assuming the standpoint of reason in Kant's argument. Consider, for example, the sentence that precedes the cited passage which begins, "Now, one cannot possibly think of a reason that would consciously receive direction from any other quarter [...] [Nun kann man sich unmöglich eine Vernunft denken, die mit ihrem eigenen Bewusstsein in Ansehung ihrer Urteile anderwärtsher eine Lenkung empfinge]." Here Kant starts off by considering how anyone must think of reason, but the 'consciously [Bewusstsein]' reveals that he does not inhabit this standpoint for long. Kant's point is not that reason cannot be imagined as determined by alien influences, but that conscious reason cannot be so imagined. But, in order to understand the properties of conscious reason, we must ask how reason views itself. Again Kant seems to suggest that we must stop asking how reason is (how it must be regarded) and asks instead how it must regard itself.

After section 2, this sort of language which references the first-person standpoint becomes a constant feature of Kant’s manner of expression. Whereas before section 2, Kant spoke directly about wills, laws, freedom, imperatives, and even reason, after section 2, Kant emphasizes that many of his claims are only true from the standpoint of the rational agent herself. Thus, at GMS: 452 "a rational being must regard himself [...] [[...]] muss ein vernünftiges Wesen sich selbst [...] ansehen]" or at GMS: 453 "a rational being counts himself [...] [Das vernünftige Wesen zählt sich [...] ]" or at GMS: 455 "all human beings think of themselves as having a free will [...] [Alle Menschen denken sich dem Willen nach als frei [...] ]", etc. Any successful account of Kant’s argument in section 2 must explain this appeal to the perspective of rational agency itself.

The standard interpretation outlined above understands this shift in perspective to contain Kant's argument that we are indeed rational agents. On this reading, much emphasis is placed on the apparent fact that while we may deny the efficacy of another person’s reasons, citing causal events as the real explanation, it seems impossible to deny the efficacy of our own reasons. When Daisy herself is deciding what to do, she assumes that the reasons amongst which she is deliberating will be efficacious in bringing about an action. If Daisy knew with absolute certainty that she was going to go shopping

---

instead of going to work, then her reflections on whether it was a good idea to go shopping would count as idle speculation rather than genuine practical deliberation. From the standpoint of Daisy’s own deliberation, Daisy must regard herself as having agentive freedom. It seems, therefore that Daisy’s agency can be defended from the first person point of view.

This reading suffers from the by now familiar problem that it takes Kant to be arguing for rational agency rather than from rational agency. If the first criterion is accepted then agency is simply assumed in this section and Kant’s point cannot be to argue for what is properly understood as a premise. The effectiveness of our reasons is contained in the definition of rational agency, and rational agency itself is not here in question. But even setting aside this familiar criticism, it seems unlikely that Kant is using the point of view of practical deliberation as an argument for rational agency for the simple reason that the standpoint of practical deliberation just is the standpoint of rational agency. Such an argument would have Kant supporting the existence of rational agency by saying that it is presupposed from the standpoint of rational agency. To someone who doubts our rational causality and hence our capacity for genuine deliberation, appealing to the standpoint of practical deliberation clearly begs the question. It is only if Kant assumes that rational causality is legitimate that he can appeal to the point of view that is revealed by this rational causality in order then to argue for some further feature of this point of view, namely autonomy.

One of the important implications of the view that Kant’s argument for autonomy proceeds by adopting the point of view of rational agency is that Kant’s argument from rational agency to freedom, and consequently also his argument from rational agency to morality is synthetic. The connection between rational agency and freedom is necessary for rational agents, but not analytic. This follows from the fact that Kant must ask about how reason views itself in order to establish his connection. An analogy might be helpful. Suppose that it is perfectly possible that the duties of a loving husband are onerous, both from the point of view of someone who is not a loving husband as well as, perhaps, an objective point of view. What is impossible is that a loving husband regards his own duties as onerous. In such a case, there is no conceptual connection between being a ‘loving husband’ and having ‘non—onerous duties’. It is nonetheless a necessary feature
of loving husbands that they regard their duties as non-onerous. Similarly, for Kant, there is no conceptual connection between ‘rational agency’ and ‘freedom’; the connection exists between ‘freedom’ and ‘how rational agents must regard themselves’. If Kant’s argument is only valid from the point of view of a rational agent herself, then the connection between rational agency and freedom is not analytic.\footnote{Recently, in “The Normativity of Instrumental Reason.” In: \textit{Ethics and Practical Reason}, Cullity and Gaut eds. Oxford. 1997, 215-255, Korsgaard has defended the claim that there is an analytic connection between rationality, freedom and categorical imperatives. According to Korsgaard, “the familiar view that the instrumental principle is the only requirement of practical reason is incoherent.” (“Normativity […],” 218) Since the “familiar view” is just instrumentalism, Korsgaard’s argument amounts to an argument for the incoherence of instrumentalism (rather than just its falsity) because she thinks non-instrumental requirements can be analytically derived for any rational agent. If, as I am arguing here, freedom (and hence morality) is connected only synthetically with rationality then instrumentalism is coherent but false. Although Korsgaard’s argument is too complicated to do it justice here, I would like to point out a direct consequence of such a view. If hypothetical imperatives, right along with categorical imperatives, require unconditional practical laws then, as practical judgments, they ought to have the same status when it comes to being analytic or synthetic. Korsgaard herself resolves this issue by attributing the synthetic character of both hypothetical and categorical imperatives to their normative status. In other words, these practical judgments are synthetic because they are imperatives that place an obligation rather than principles which express a logical necessitation. Obviously, this flies in the face of Kant’s view that hypothetical imperatives are analytic and their possibility is easily understood whereas categorical imperatives are synthetic and their possibility requires a deduction. These issues are treated more thoroughly in the author’s forthcoming article “Do Hypothetical Imperatives Require Categorical Imperatives?” \textit{European Journal of Philosophy}.}

To say that section 2 involves a synthetic a priori proposition has far-reaching implications for how section 2 fits in with the rest of the \textit{Groundwork}. In particular, since Kant believes that every synthetic a priori proposition requires a transcendental deduction, the fact that section 2 is synthetic naturally raises the possibility that the deduction of which Kant speaks at many points in \textit{Groundwork III} (GMS: 447, 454, 463) actually takes place in section 2. At the end of section 1, Kant says that some “further preparation is required [\textit{noch einiger Vorbereitung}]” (GMS: 447) implying that section 2 (and possibly section 3 as well) is preparation for a deduction rather than a deduction.
itself.\textsuperscript{15} However, the fact that section 2 contains a synthetic step suggests that this preparation represents an important, if not the important, step within the deduction rather than simply necessary preliminaries. Perhaps, section 2 ought to be thought of as the central lemma and the deduction itself a mere corollary, or perhaps section 2 ought to be thought of as providing a preliminary overview of the deduction which the later sections fill in by responding to objections. In any case, the synthetic step in section 2 places it at the center of any understanding of the deduction in \textit{Groundwork} III.

C. Kant’s argument is practical all the way through

The final feature of section 2 is that the argument should be understood to be about practical reason from start to finish. It is not an argument directed at theoretical reason, nor does it argue by analogy to theoretical reason. Although Kant himself does not mention theoretical reason or its relation to practical reason in the second section, it is widely assumed that Kant’s argument depends upon an analogy from theoretical reason to practical reason.\textsuperscript{16} There seems to be two main reasons for making this assumption.\textsuperscript{17}

\begin{itemize}
\item \textsuperscript{15} Schönecker, “How is a categorical Imperative Possible?” 302, and Timmerman \textit{Kant’s Groundwork} [...], 126 have argued that the deduction in \textit{Groundwork} III must take place in section 4, and that the “Preparatory argument” (the term belongs originally to Allison, Henry: “Kant’s Preparatory Argument in Grundlegung III”. In: \textit{Grundlegung zur Metaphysik der Sitten : Ein Kooperativer Kommentar}, Otfried Höffe, ed. Frankfurt am Main. 1989) must take place in both sections 2 and 3. Allison “Kant’s Preparatory Argument […]” argues that the “preparatory argument” takes place only in section 2. The synthetic nature of section 2 shows that the distinction between preparation and deduction is not as clear-cut as Schönecker and Timmerman think. Part of the preparation for an argument can be a step within the argument. Indeed, it could be the crucial step.
\item \textsuperscript{16} Ameriks, “Kant’s Groundwork Argument […].” 228. Henrich, “The Deduction […],” 312. Paton, \textit{The Categorical Imperative}, 218f. Wood, \textit{Kant’s Ethical Theory}, 177. Schoenecker, \textit{Kant: Grundlegung III}, 208ff. Of these only Schönecker defends this interpretation in any detail. His defense rests solely, however, on an appeal to the “Review of Shulz.” “So könnte man diese Stelle [i.e. the last 3 sentences of section 2] womöglich lesen [i.e. to be exclusively about practical reason rather than theoretical reason]-gäbe es nicht eine Parallelstelle in der schon erwähnten ‘Recension von Schulz’ […]” (225). Only Mieth and Rosenthal, “ ‘Freedom must be presupposed […]’ take Kant to be arguing directly from practical reason to autonomy.
\item \textsuperscript{17} I think the implicit third reason is that, most interpreters believe that only a theoretical assumption could count as an independent argument for morality. Since they assume that
The first reason is that, although Kant does not explicitly mention theoretical reason, his language seems to be more appropriate to a theoretical sphere. Thus, he talks of the “judgments [Urteile]”\(^{18}\) of reason and uses phrases like “reason regarding itself as the author of its own principles”. Theoretical reason is most naturally thought of as making judgments, and theoretical reason is more naturally spoken of as ‘regarding itself’ since ‘regarding itself’ seems to be a way of talking about forming beliefs. The second reason that is cited as evidence that Kant must be talking first and foremost about theoretical reason stems from considerations that take us beyond the text of the *Groundwork*. Nevertheless, because these considerations are so widespread,\(^ {19}\) I would like to consider them here. At approximately the same time as the *Groundwork*, Kant wrote a review of the first part of Schulz’s work, in which Schulz defended an extreme form of fatalism that denies to us both theoretical and practical freedom. In this review, Kant criticizes Schulz’s fatalism seemingly on the grounds that theoretical reason must regard itself as the author of its own principles. For a fatalist to believe that both ‘I believe fatalism to be true because it has the strongest reasons supporting it’ and to believe that ‘I only believe fatalism to be true because of my causal circumstances’ is a contradiction.\(^ {20}\) Crucially, Kant then argues that “just in the same way [ebenso]”\(^ {21}\)(RezSchulz: 14) practical reason must also regard itself as free. This suggests that in the ‘Review of Schulz,’ at least, Kant argues for the necessity of freedom by analogy to the impossibility of theoretical reason regarding itself as determined. Hence, commentators conclude, the *Groundwork* must contain a similar argument.

Consider first the argument for the introduction of theoretical reason based on the “Review of Schulz”. First of all, it should be pointed out that Schulz’s strong brand of fatalism rules out not only autonomy but also all forms of rational causality. Schulz

\(^{18}\) Thus, Paton, *The Categorical Imperative*, 218, argues that reference to ‘judgments’ alone is sufficient to justify his belief that the argument takes a “surprising turn” by appealing to theoretical reason.


doubts not only that we are capable of acting on non-instrumental reasons, but also that we are capable of acting on reasons at all. If the first criterion for interpreting section 2 is accepted, then in section 2 of *Groundwork* III, Kant has simply assumed the rational causality that is very much the central question in the “Review of Schulz”. With such vastly different intentions, it is dangerous to use the review as a clue on how to interpret the *Groundwork*. But even if my first criterion is not accepted, I still believe that the evidence that the “Review of Schulz” provides is inconclusive. In the review, Kant does make a distinction between theoretical and practical reason, and he does argue that a fatalist contradicts his own fatalist beliefs when he affirms that fatalism is correct. However, his argument that the fatalist must *also attribute* to himself “freedom of the will in acting” turns out to be an *ad hominem* attack directed specifically at Schulz, that assumes personal knowledge of Schulz’s commitment to “righteous conduct [rechtschaffen Lebenswandl]”:

In the same way he [Schulz] must also assume freedom of the will in acting, without which there would be no morals, when – as I’ve no doubt – he wants to proceed in his righteous conduct in conformity with the eternal laws of duty and not to be a plaything of his instincts and inclinations. (RezSchulz: 14)

Not only does Kant assume knowledge of Schulz’s character (Schulz’s good intentions are praised throughout the essay), but Kant seems also to assume that freedom of the will is only necessary for those who *want* to proceed in their righteous conduct. Rather than being an independent argument for practical freedom based on an analogy with theoretical freedom, what we get in the ‘Review of Schulz’ is an argument that claims that freedom is necessary for those who are committed to acting morally. Given this, and given the fact that Kant’s target in *Groundwork* III is not the fatalist who appears in the review, it seems very questionable to take the “Review of Schulz” as evidence that Kant’s argument in *Groundwork* III proceeds by analogy to theoretical reason.

---

21 “*Ebenso muss er auch Freiheit des Willens im Handeln voraussetzen, ohne welche es keine Sitten gibt, wenn er in seinem, wie ich nicht zweifle, rechtschaffenen Lebenswandel den ewigen Gesetzen der Pflicht gemäss verfahren und nicht ein Spiel seiner Instinkte und Neigungen sein will [...]”
As for the claim that Kant’s language is more appropriate in the theoretical sphere, the supposed inappropriateness arose from two facts, neither of which is very persuasive. The exclusive right of theoretical reason to the term ‘judgment’ is explicitly denied in the *Critique of Practical Reason* in which Kant talks of a pure practical faculty of judgment and compares it with the theoretical variety.22 Nor can much weight be placed on the fact that only theoretical reason can be thought of as regarding itself in a certain way. Surely beliefs can be embedded in practical deliberations and someone who deliberates in the light of those beliefs could be said to practically regard herself in a certain way. This regard could be independent of anything that she theoretically believes. One might find oneself consistently making plans for the contingency that one is going to fail. A belief in our failure is thus orienting our practical reasoning, but it may very well have little to with our theoretical assessments of the probability of our success. A belief about how the world is can be practically orienting as well as theoretically orienting. Neither of these two facts, therefore is sufficient to warrant the near universal assumption that Kant has suddenly introduced theoretical considerations into what is explicitly a practical argument.

Taking all three features at once we get the following: a successful interpretation of the second section of *Groundwork* III must 1) start with rational agents and end with agents capable of acting on non-instrumental reasons, 2) adopt, at some point, the point of view of the rational agent herself, and argue from within this point of view that it is necessarily possible for us to act on non-instrumental reasons, and 3) be framed entirely in terms of practical reason, and must not argue first from theoretical reason and then only by analogy to practical reason. Such an argument would depend crucially upon an understanding of what it would mean for practical reason to regard itself as incapable of acting on non-instrumental reasons.

II. A speculative sketch of Kant’s argument in section 2 of *Groundwork* III

Let us start then, as the first feature instructs us to, with an instrumentally rational agent – a creature who has the capacity to act for reasons but may not have the capacity

22 KpV: 67f.
to act for non-instrumental reasons. The first point to make is that according to the instrumentalist picture that we have supposed, instrumental reasoning itself is something that is incapable of practical evaluation. Hypothetical imperatives instruct us to adopt the means to our specific ends. They tell us to buckle up if we want to avoid injury and to back up our hard drive if we want to avoid data loss. But, instrumental reasoning itself is not up for practical evaluation. Thus, if we ask an instrumentalist why we ought to back up our data, she will say that it is to avoid data loss. But if we ask the instrumentalist, why we should in general adopt the means to our ends, she is liable to think the question nonsense. According to the instrumentalist picture, means-ends reasons are the only sorts of reasons there are. To demand a reason for means-end reasoning itself can only be understood as a demand that we supply a desire that means-end reasoning is an instrument towards. ‘Why should I adopt the means to my ends?’ can only mean, ‘Why is following means-end reasoning in my interest?’

But, if this is plausible, then an instrumentalist cannot account for a certain type of practical evaluation of means-end reasoning itself. She can’t account for us taking means-end reasoning itself to be good in a non-instrumental way. This might seem unsurprising. How could an instrumentalist approve of instrumentalism in any way other than as a means to a desire that we have? But the upshot of this is that there is a certain mode of practical self-evaluation that is cut off from instrumental evaluation. An instrumentalist has a way of judging of everything else whether it is good or bad (in terms of her desires and beliefs) but means-end reasoning itself seems to call for some sort of evaluation different from this. If we are persuaded that questions about the practical validity of instrumental reason are legitimate (and, I admit this is a big ‘if’), then we have adopted a perspective from which instrumental reasons cannot be the only form of practical evaluation. We have transcended means-end reasoning. But why should we think that such questions are legitimate?

It is here that I think Kant invokes the second and third features. Instead of asking how rational agents are, Kant asks how rational agents must practically (feature three) regard themselves (feature two). Kant’s argument rests on the assumption that forming a practical regard about one’s own agency is to become practically self-conscious of instrumental reason, which is itself to raise questions about the goodness of instrumental
reason. This assumption is plausible. To become theoretically conscious of something is to ask of that thing whether it is true. To become practically conscious of something is to ask whether it is good. To become practically conscious of instrumental reasoning, therefore, is precisely to raise questions about the goodness of instrumental reasoning. Contemplating rational agency as a rational agent is a form of self-consciousness, and to become self-conscious of our own reasoning from a practical point of view is precisely to raise questions about this instrumental agency that cannot be answered from within an instrumentalist picture. Instrumentalism is compatible with heteronomy, self-conscious instrumentalism, on the other hand, is not. If such an argument is at all plausible, it has started with instrumental rational agency, proceeded by adopting the point of view of a practically rational agent, and then, by applying this point of view to instrumental agency itself, it has concluded that we are capable of acting on non-instrumental reasons.

Conclusion

In this paper, I have argued that a successful interpretation of section 2 will have 3 features, and I have put forth a tentative argument, which is consistent with all three features. While I hope to have shed some light on section 2, I have said nothing about the notorious “suspicion of the circle [Zirkelverdacht]” that follows section 2 in which Kant seems to disavow the conclusions of section 2. Although I cannot, here, give a full answer to this problem, I would like to conclude by briefly sketching how my interpretation of section 2 bears on this issue. According to my interpretation, section 2 contains an argument that starts with rational agency and proceeds synthetically to autonomy. Since this argument is both synthetic and a priori, it represents a key step in the deduction of the moral law, if it is not equivalent to the deduction itself. Why then does Kant seemingly disavow the results of his deduction in section 3 when he claims that the argument thus far must be “freely admitted [frei gestehen]” to proceed in a “kind of circle [Art von Zirkel]”? What hope do we have of defending and reconstructing an argument that Kant himself immediately concedes is circular? Although there are many possible responses to this, I would like briefly to propose a response that I think is compatible and in the spirit of the main theses of this paper. I have argued that any
rational agent that becomes practically self-conscious of the form of her practical reason must also regard herself as autonomous. But I have also freely admitted that this argument depends upon the assumption that raising practical questions about our own agency is a sensible and legitimate thing to do. The argument was of the form, ‘If a certain type of question makes sense, then only the assumption of autonomy would legitimate such questions’. But what of someone who refuses to grant sense to this type of question. What of a hard headed instrumentalist who refuses to consider any questions about the legitimacy of instrumental reason that cannot be turned into questions about what instrumental reasoning is good for? Such a hardheaded instrumentalist might wonder whether I ought to take up a position of practical self-consciousness about the agency itself. The resolution of the circle, I suggest, represents Kant’s answer to this question.