

Kant on Misology and the Natural Dialectic

John J. Callanan

King's College London

© 2019 John J. Callanan

*This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivatives 3.0 License.
<www.philosophersimprint.org/019047/>*

1. Philosophy and Common Moral Cognition

Towards the conclusion of the First Section of the *Groundwork of the Metaphysics of Morals*, Kant describes a process whereby a subject can undergo a certain kind of moral corruption. This process, which he calls a “natural dialectic”, can cause one to undermine one’s own ordinary grasp of the demands of morality (4: 405).¹ The threat of the natural dialectic is of particular interest since it not only gives a precise causal account of the phenomenon of moral corruption but also constitutes Kant’s case for the relevance of philosophy to everyday practical life. The question as to whether and how philosophy can have any practical significance within our ordinary moral lives is one that Kant himself had just raised at the conclusion of the section. According to his own account of “common moral cognition”, the cognitive capacity of ordinary human beings “is very well informed in all cases that occur, to distinguish what is good, what is evil, what conforms with duty or is contrary to it” (4: 404).

This commitment stemmed from Kant’s reading of Rousseau in the 1760s.² Rousseau had convinced him that the unreflective responses of ordinary uneducated human subjects are more reliable than those of philosophical experts. Kant’s previous prioritization of the improvement of the intellect and the thought that “this alone could constitute the honor of mankind” later struck him as constituting both a philosophical and personal failing. Kant famously confessed in the notes to

1. References to the *Groundwork* are to (Kant 1786/2011). References to Kant’s other writings are to the Cambridge Edition series. References to Kant’s writings in general are to the *Akademie* German edition of Kant’s works. Abbreviations used are as follows:

(A/B) = *Critique of Pure Reason*
(Anthropology) = *Lectures on Anthropology*
(Corr.) = *Correspondence*
(Metaphysics) = *Lectures on Metaphysics*
(Notes) = *Notes and Fragments*
(Practical Reason) = *Critique of Practical Reason*
(Logic) = *Lectures on Logic*

2. 4: 404. For the positive influence of Rousseau, see (Ameriks 2012a; Cassirer 1983; Henrich 1992; Shell 2009; Velkley 1989; Zammito 2002). I don’t address the issue of the proper characterization of the epistemology of common moral cognition here.

the *Observations* that it was Rousseau who had set him straight on his previous “blinding superiority” and instead affirmed the theoretical centrality of the moral responses manifested by ordinary agents.³ The Kant of the Critical period retains this commitment (to what I will call *Rousseau’s premise*) in holding that the primary reliable data for moral philosophy ought to be the immediate responses of ordinary people. In the *Critique of Pure Reason*, Kant states early on that “without doubt the concept of **right** that is used by the healthy understanding contains the very same things that the most subtle speculation can evolve out of it” (A43/B61).⁴ The challenge is to give a plausible characterization of how this could be so, of how people’s ordinary responses might be *rational* in character, thereby grounding the claim that ‘the voice of reason in reference to the will [is] so distinct, so irrepressible, and so audible even to the most common human beings’ (*Practical Reason*, 5: 25).⁵

Kant’s aim is to show that “human reason, even in the commonest understanding, can easily be brought to a high measure of correctness and accuracy in moral matters” (4: 391). Given the initial assumed

reliability, however, it is sensible to ask whether or why philosophy is even needed to bring about this high measure of correctness. It is reasonable to wonder whether we might perhaps eschew philosophy altogether. It seems that our ordinary capacities simply don’t require “science and philosophy” for moral guidance. As such, it might be better to leave the management of our lives to the mostly unreflective first-order exercise of our moral capacities.⁶ If common moral cognition is sufficient, then surely philosophy cannot be necessary. If anything, philosophy’s influence constitutes a potential threat, since one must now make sure that philosophy does not in fact “lead common human understanding away from its fortunate simplicity” (4: 404). Rousseau’s premise regarding the first-order reliability of moral capacities thereby problematizes the issue of the practical relevance of philosophy.

Kant recognizes this problem and responds by claiming that despite common reason being perfectly sufficient for first-order moral guidance, the natural dialectic *nevertheless* generates a distinct kind of threat that entails that second-order philosophical inquiry is necessary. The nature of that threat is such that the reliability of the first-order judgments issued by common moral cognition is insufficient protection from a peculiar kind of moral corruption.⁷ Kant’s claim here raises several questions. Firstly, what kind of sufficiency is common moral cognition really supposed to have if it nevertheless allows for moral corruption? Secondly, why did Kant think that *philosophical* inquiry in particular is necessitated as a response to this threat (rather than simply more determined first-order moral instruction, for example)? Thirdly, how might philosophical inquiry then subsequently suffice as a response to that threat? The explicit goal of the *Groundwork* is “the identification and corroboration of the supreme principle of morality” (4: 302), but there must be a coherent narrative about how that goal,

3. (Kant 2005, 7, 2:216–17 [Ri 37–9]) — see (Neuhouser 2008, 112). Shell and Velkley call this encounter with Rousseau Kant’s “philosophic rebirth” (Shell and Velkley 2017, 193).

4. See also 4: 389, 4: 394, 4: 412, 4: 454. For Kant’s repeated general references to the importance of respecting common cognition in both the theoretical and practical spheres see Bxiv, B3–5 A184–B227, A358, A473–4/B501–2, A480/B508, A839/B851, *Practical Reason* 5: 36, 5: 70, 5: 91–2, *Notes* 16: 374. For a discussion of this methodological commitment, see (Callanan 2019). The claim in the *Groundwork* is not of course that of presupposing the truth of common moral cognition’s claim, but rather only the conditional methodological constraint that *if* a supreme principle of morality is possible, then it must be one such as would explain the reliability of common moral cognition.

5. As will be discussed, one of the challenges of such a characterization is to present a picture of our rational faculties such that they can operate in producing immediate and non-reflective feelings possessing motivational efficacy, a characterization traditionally presumed antithetical to identifying such mental states as rational in nature. Kant’s account of respect [*Achtung*] is just such an attempt to explicate “grounds of motivation that, as such, are represented completely a priori by reason alone ...” (4: 391). However, I do not detail or evaluate that account in this paper.

6. For this point, see (Thorpe 2006; Timmermann 2007b).

7. I characterize “moral corruption” in the following section.

even if realized, might integrate with the initial claim of the sufficiency of common moral cognition.⁸

Kant says that philosophy is required by common moral cognition “not in order to learn from it, but to obtain access and durability for its prescription” (4: 405). He seems to hold then that what is added by philosophical inquiry is not some new set of first-order moral prescriptions. Common reason does not “learn” anything in this way.⁹ Rather, philosophical inquiry is thought to make those first-order prescriptions more “durable” in an agent’s mind. What this just might mean, however, is unclear. One might push again the objection that if common reason cannot offer “durable” first-order moral guidance, then it

8. To clarify: There is an obvious sense in which establishing the possibility of moral value might reinforce a disposition to judge morally that is already in place. There is also an obvious sense in which detailing the categorical imperative procedure might aid a subject in their practical deliberation. Kant’s claim, however, is a stronger one than that it is *possible* that philosophy can help ordinary cognition in these ways. He claims that there is a threat of moral corruption, one that is “natural” to educated and uneducated human beings alike, such that philosophical reflection is *necessary* for addressing this threat. It is the claimed necessity of philosophy to practical life that is the focus of this paper.

9. Compare Kant’s mockery of a reviewer of the *Groundwork* who complained of the lack of any new moral principles put forward in it:

But who would even want to introduce a new principle of all morality and, as it were, first invent it? Just as if, before him, the world had been ignorant of what duty is or in thoroughgoing error about it. (*Practical*, 5: 8, *note*)

Kant’s aims in the *Groundwork* are not to instill an interest in morality where there previously was none. Rather, he asserts that “[t]he human mind takes (as I believe is necessarily the case with every rational being) a natural interest in morality” (A829–30/B857–8, *note*). One might think of the use of the categorical imperative procedure as a kind of non-moral algorithm to which any agent might appeal in order to generate moral commitments in various scenarios. For what I regard as compelling opposition to this familiar picture, see (Geiger 2010). In what follows, it should become clear that I don’t take Kant’s account of philosophizing to be that of revealing to ordinary moral agents new moral truths of which they were hitherto ignorant. This is required, I would claim, by Kant’s claim that the *Groundwork* concludes by bringing one “back to common cognition” (4: 392) and that the deduction of freedom is “confirmed” by common cognition (4: 454). For discussion of the latter point, see (Sticker 2014).

is simply not the case that common reason is competent to distinguish “what conforms with duty or is contrary to it”. Conversely, if common reason *can* do this, then it is hard to see what philosophy might be in an exclusive position to add.

This raises the worry that the function of the natural dialectic passage is perhaps merely structural (and to that extent, artificial). Kant orders the three parts of the *Groundwork* around required “transitions”, beginning in the First Section with an initial analysis of ordinary moral psychology. He then attempts to “transition” from that analysis to the more convoluted and overtly technical reflections of the Second Section. The aim of the First Section is explicitly expressed in its title: “Transition from common to philosophical moral rational cognition” (4: 393). As such, he requires some justification for that transition. Yet since an essential part of Kant’s initial claim about common cognition is its sufficiency, Kant is faced with a challenge in explaining the necessity of any transition at all. One might think then that Kant’s concluding presentation of natural dialectic satisfies that function in a somewhat *ad hoc* manner.

My primary thesis is that the natural dialectic is in fact entirely central to any understanding of Kant’s philosophical project in the *Groundwork*. This is a significant claim, though attention to the specific historical context of Kant’s rhetoric in the First Section substantiates it. Consideration of the historical context can also reveal Kant’s proposed answers to the three questions raised above. Such consideration clearly reveals that the natural dialectic passage stems again from the influence of Rousseau. No doubt because of the aforementioned commitment to common moral cognition, the influence of Rousseau upon Kant’s intellectual development is more often than not presented in a positive register.¹⁰ However, I would claim that this undeniable fact

10. E.g. (Alberg 2015; Ameriks 2012a, 2012b; Hohenegger 2012; Quadrio 2009; Velkley 2013, 1989). Kant’s points of resistance to Rousseau are valuably noted in all of the above, as well as in (Cassirer 1963; Shell 2009; Shell and Velkley 2017). Perhaps the most commented upon aspect of Rousseau’s influence is with regard to autonomy, e.g. see (Zammito 2002) and various papers in (Sensen 2012).

regarding Rousseau's positive influence has led commentators to neglect the quite specific ways in which passages of the *Groundwork* are directed *against* Rousseau.

This opposition to Rousseau was expressed in specific passages where that opposition would have been immediately grasped by Kant's intended audience. That these passages are relatively passed over by contemporary readers has the consequence that their centrality to the *Groundwork's* aims is missed.¹¹ Rousseau's arguments in the *First Discourse* were widely seen not just as claiming that philosophy is unnecessary for moral guidance, but that philosophy in fact directly undermines our moral capacities. In the *First Discourse*, Rousseau infamously attacked the value of scientific — including philosophical — reasoning to human flourishing in general. Moreover, while in the *First Discourse*, Rousseau presented philosophy as pernicious in broad cultural terms, in *Émile*, Rousseau presented the phenomenon of the self-undermining of one's recognized moral commitments as the first-personal manifestation of that same phenomenon. Kant's raising of the thought that philosophy might lead the common understanding away from its "fortunate simplicity" is an explicit reference to this Rousseauian threat.

Kant saw an opportunity to address this threat through the discussion of the natural dialectic.¹² There has been surprisingly little analysis of the mechanism of the natural dialectic.¹³ My aim is that a

11. An exception is Shell, who is particularly sensitive to the centrality of the natural dialectic to Kant's thought — see (Shell 2009, 151, 255).

12. Some care is needed, however, to distinguish the somewhat caricatured Rousseauian picture with which Kant is concerned from Rousseau's own more nuanced views. References to "Rousseau" throughout refer to the caricatured image usually formed by his popular readership.

13. Recent Anglophone summary accounts can be found in (Allison 2011; Schönecker and Wood 2015; Timmermann 2007a). (Muchnik 2010) notes the importance of the natural dialectic passage and connects it to Kant's account of radical evil later in the *Religion*. Grenberg briefly mentions the possibility of a Rousseauian context, though does not regard it as central to Kant's account of corruption in the *Groundwork*, on the grounds that the move to corruption occurs on "individual rather than societal lines" (Grenberg 2013, 87, fn 6.). My argument here is that the Rousseauian critique attempts to parallel the

reconstruction of natural dialectic process reveals that it relates to a quite specific worry, namely that ordinary agents can come to the belief that some moral demands are not such that they *always* override other non-moral concerns.¹⁴ That the phenomenon arises at all is, I claim, intimately related to a different concern, that of *misology* or the hatred of reason (4: 395).¹⁵ Kant's aim is to refute one who might have doubted the rational authority of morality's peculiar insistence upon the exceptionless universality of its prescriptions.¹⁶ Given Kant's view that the demands of morality *are* in fact categorical and thus strictly universal, the natural dialectic is his account of how ordinary agents could come to deceive themselves as to both the extent and nature of their obligations. His analysis is that they come to do so by obscuring from themselves the rational character of their obligations. The task of the philosophy in general, and the *Groundwork* in particular, is to afford the means for such a subject to remove that particular unclarity from their own consciousness.

While Kant's discussion of misology is determined by Rousseau's general appropriation of 18th-century debate on the nature of societal corruption, the natural dialectic passage is targeting a very specific passage in *Émile*, one where Rousseau presents his own account of interior moral struggle. Here, Rousseau brings the suspicion of reason from the cultural context of the arts and sciences to that of first-personal moral

societal and individual threat of corruption in the *First Discourse* and *Émile* respectively, and that Kant's response is sensitive to just this parallel.

14. The challenge I am considering is from one who might even concede that morality is real yet asks what grounds we have to prioritize its dictates over other interests in every circumstance — for discussion, see (Grenberg 2013).

15. The reference is to *Phaedo* 89d–e (Plato 2002, 127). I return to this allusion in §5.

16. Kant would already have been familiar with this type of position: For a single obvious widely-read example, Bayle's characterization of Pyrrhonism was one whereby it held that moral obligations held at least a customary default warrant for subjects. The Pyrrhonist only doubted obligation's rational basis, and for that reason suspended judgment "on the question of whether such and such an obligation is naturally and absolutely legitimate; but they did not suspend judgment on the question of whether it ought to be fulfilled on such and such occasions" (Bayle 1991, 195, "Pyrrho").

deliberation. The diagnosis concerns two questions: firstly, whether the cause of the corruption is natural or societal; secondly, whether philosophy is better thought of as restraining or exacerbating that corruption. In the *First* and *Second Discourses*, Rousseau had maintained that society was the occasioning cause of moral corruption and that philosophy only made the situation worse. Kant maintains that human nature itself is the cause of moral corruption, but that philosophy can and must be used to address this existential predicament.¹⁷ Moreover, Kant's ingenious characterization of the mechanism of corruption has it stem from the very fact that our ordinary moral demands are categorical in character.

The *Groundwork* abounds in argument, arguments that have been nearly exhaustively examined. Comparatively little attention, however, has been paid to the *rhetorical* dimension of the *Groundwork*, that is, to the non-argumentative but suasive role certain passages are supposed to have upon the reader. Such passages, however, set the scope for what is at stake in the accompanying arguments. By attending solely to argument and neglecting the rhetorical context, we achieve at best a caricature of positions in the history of philosophy. What is at stake in both the misology and natural dialectic passages is the framework of human nature chosen to interpret a certain class of phenomena. These phenomena are those relatively invariant, immediate, and non-reflective responses to morally salient scenarios that ordinary subjects seem to issue. For Rousseau, it was obvious that these features render them *non-rational* responses. Neglect of this obvious fact was required, Rousseau thought, in order to sustain the Enlightenment's elitist fetishization of rationality within modern culture.¹⁸ Moreover, the particular phenomenon of first-personal moral evasion constituted an

individual-level example of the enervating effect of the cultural prioritization of rationality. Kant's rhetorical aim is to present an interpretation of the data whereby the situation is precisely reversed: the antagonist in cases of moral evasion is inclination; the protagonist is rationality. If this reconstruction is accepted, then Rousseau's characterizations can be recast as symptomatic of an unwarranted demonization of rationality and concomitant fetishization of non-rational feeling within 18th-century culture.

The rest of the paper is structured as follows. In §2, I set out the different senses in which an agent can become unresponsive to the demands of morality. In §3, I outline in broad terms the challenge to philosophy and the model of moral corruption presented by Rousseau. In §4, I reconstruct Kant's analysis of the misologist's argument against rationality. In §5, I discuss how Kant's account of misology was formed for the purpose of reclaiming the figure of Socrates against Rousseau. In §6, I finally turn to a reconstruction of the natural dialectic passage. Kant's remarks on the matter are brief and few commentators provide an analysis of either the conditions under which the natural dialectic arises or of the exact details of its operation. Kant has an imaginative and psychologically sensitive characterization of the process of moral self-deception worthy of re-examination. I conclude in §7 showing that the passage responds directly to a similar one in *Émile* by characterizing the phenomenon of moral corruption as one that does not undermine, but rather reinforces the need for the philosophical establishment of a "culture of reason" (Bxxx, A850–1/B878–9). The account of moral corruption and the demand for philosophical redress thus goes to the heart of not just the *Groundwork*, but of Kant's initial presentation of the Critical project generally.

2. The Threat of Moral Corruption

Why did Kant write the *Groundwork*? Some familiar answers are that he wrote it to establish ethics as a science (4: 387), to secure the supreme principle of morality (4: 392), and to banish the thought that morality is chimerical (4: 407, 4: 445). These answers — all true — can

17. This claim must be qualified by Kant's mature Critical thoughts regarding the development of reason within society. My claims throughout this paper are restricted to the early Critical presentation of reason in the *Groundwork*, and specifically that the titling of the dialectic as "natural", must be understood as a deliberately un-nuanced rhetorical strategy made to highlight Kant's anti-Rousseauian agenda to his readers.

18. For general discussion, see (Beiser 2017).

give rise to the impression that Kant thought that by securing ethics as a science he might thereby combat moral scepticism in ordinary life. However, it is clear that Kant's focus upon common moral cognition, understood as a widespread assumption of the reality of moral dispositions, complicates this picture. Kant did not think that moral scepticism was a widespread view among ordinary people yet he *did* think that the philosophical reflections contained in the *Groundwork* were necessary for something that would nevertheless be decidedly for their benefit.

Kant's approach in the First Section makes distinct appeals to the phenomenology of everyday moral life, in the sense that it proceeds from observations on the particular features present to ordinary consciousness when one takes oneself to be moral responsive.¹⁹ He points to some general felt features of our moral responses, such as their sense of necessitation and of their apparent universality.²⁰ Given that we do feel such concerns, and given that — if Kant's analysis is correct — this feeling is really a responsiveness to the universality of certain prescriptions, then one can articulate those feelings in general by appeal to a simple expression of the universal law formulation, considering whether one's proposed course of action might hold for everyone without incoherence (4: 402). It seems that Kant holds that as far as practical guidance is concerned, some brief and not particularly philosophical reflections such as these are sufficient.

In fact, Kant starts from a point whereby he takes it for granted that when we ask a question as to whether an action is right, we take it that a positive answer to this question entails attributing the rightness in question a higher status within our practical deliberations over considerations of that action's advantageousness, etc. There is then a recognized fact that our moral demands speak to us with an authoritative voice within our practical deliberations.²¹ However, one can grant

this claim about the phenomenology of moral considerations and still ask why we should think that this apparently authoritative voice is in fact *always* authoritative. This question is one that is neither asked nor answered by common moral cognition. Kant indicates as much when he says that just by registering the authority of respect for the moral law "I do not yet see [*einsehe*] on what it is founded (which the philosopher may investigate)" (4: 403).²² This question, the question of explanatory insight [*Einsicht*] into the *authority* of reason, he claims, is a distinctively philosophical one.

That one can act consistently out of respect for the law while manifesting this kind epistemic deficit already might be thought to put the subject at some risk. Before proceeding further, it is worth clarifying just what particular phenomenon Kant is concerned with in the natural dialectic passage. It is best understood in contrast to another theme in Kant scholarship, which concerns the sense (if any) in which Kant is engaged with moral scepticism.²³ For example, it has recently been argued that Kant is not in the business of providing a reason to be moral.²⁴ It is reasonably inferred from this that he is therefore not aiming to refute a moral sceptic. Others argue that, while conceding the point that Kant was not providing a reason to be moral, Kant can nevertheless be thought of as engaged with a sceptic of some type.²⁵ It would be

19. See (Grenberg 2013).

20. E.g. 4: 389, 4: 400, 4: 401, *note*, 4: 405.

21. See (Brink 1997) for discussion.

22. I take it that by talking of *Einsicht*, Kant has in mind a profound kind of understanding. The matter is of course complicated by the *Groundwork's* ultimate conclusion that just this level of understanding is impossible (4: 463). I do not explore here the forms of moral understanding provided (or of those denied) by the *Groundwork's* account.

23. Of course, there are a wide variety of sceptical positions with regard to morality that I don't canvas here. For a sample discussion see (Copp 1991; Harman 1977; Sinnott-Armstrong 2006).

24. See (Allison 2011; Hill Jr 1985; Stern 2010; Thorpe 2006; Timmermann 2007a; Wood 2008). In this regard, they are opposing Prichard's original contestation that Kant is (mistakenly) attempting to offer a reason to moral — see (Prichard 1912).

25. (Guyer 2008; Ware 2014). I agree that Kant is not interested in providing an agent who sees no initial reason to engage in moral enterprises with a reason to engage in them. There nevertheless remains at least one real sense in which Kant is interested in providing the agent with a reason to be moral. In

natural to think then that the natural dialectic passage concerns Kant's later ambition to defeat *this* sceptical challenge. If this were the case, then the natural dialectic is a process whereby upon reflection, one comes to think of morality as a phantasm.

This is not how the natural dialectic is presented. It is not contentious to claim that the natural dialectic describes a process whereby, roughly speaking, a subject become unresponsive to the demands of morality. There are different ways in which this might occur, however. One might adopt a self-consciously sceptical pose and declare morality to be a phantom of the brain. On the other hand, it could be that the bother of following one's moral obligations can subconsciously prompt one to reason oneself out of the particular obligations that one initially recognizes in the ordinary course of life. On occasions such as these, one doesn't decry morality *überhaupt*, but rather tries to argue to oneself that an initially recognized and bothersome obligation might just not apply in this particular instance. On such occasions, one does not become a moral sceptic, though one does become morally *corrupted*.

Detailing these differences cases is worthwhile, since I will argue that Kant's account of the natural dialectic — and with it, the justification for moral philosophy itself — is targeted on the problem of moral corruption rather than moral scepticism. Kant views this type of morally unresponsive attitude as pernicious just because it allows one to challenge the reality of morality without realizing that one is doing so. The effect is arguably more pernicious than moral scepticism just because the degradation of our moral commitments is undergone while we nevertheless pay lip service to the idea of respecting the demands of morality. Crucially perhaps, moral corruption is also plausibly a far more common real-world phenomenon than that of decrying morality

the first place, Kant is offering a reason for the agent to think that morality is not "the mere phantasm of a human imagination overreaching itself through self-conceit" (4: 407). If one *did* think that this were the case, then one would have a *pro tanto* reason to disregard morality's demands. In arguing that this is not the case, Kant is at least providing the agent with a reason to regard one's initial interest in morality as a genuine one.

per se. The primary motive for Kant's discussion of the natural dialectic is to afford a different explanation of the widespread phenomenon of self-incurred moral failure than the one offered by Rousseau, who had placed reason in the role of antagonist in this process of self-deception.

3. Rousseau and the Luxury of Philosophy

To see the relevance of Rousseau to this analysis, however, the natural dialectic passage must be related to a different one in the First Section. This latter passage concerns Kant's discussion of the threat of "misology". In order to appreciate the sense in which Rousseau is the target here, some brief recapitulation of the familiar themes of the *First Discourse* is required.²⁶ There are three especially relevant themes: firstly, the account of the origin of moral corruption; secondly, the account of the role of philosophy and the motive for its pursuit; thirdly, the account of rationality and its cultivation in for the human condition. As is well known, Rousseau argues that the sciences and the arts in modern society — contrary to expectations perhaps — "ha[ve] added nothing to our genuine felicity" and have led to the corruption and degradation of morals' (Rousseau 1750/2008, 26, OC III, 28).²⁷ The "men who make up the herd that is called society" have by necessity come to behave guided by custom, in deference to laws of politeness and invented models of propriety (Rousseau 1750/2008, 8, OC III, 8). The outcome of the move to cultivated enlightenment in society has

26. A standard account of the *First Discourse* can be found in (Armstrong Kelly 2001; Dent 2006, Ch. 3).

27. Throughout this section, I will be presenting a rather un-nuanced view of Rousseau's position. It is my contention that Kant is more accurately thought of responding to a received caricature of Rousseau than Rousseau himself. It is the former that Kant thought particularly pernicious. It is beyond the scope of this paper to determine how many Rousseauian claims Kant attributed to Rousseau himself as well as to those who were under his sway (for an example of his more nuanced understanding, see *Anthropology*, 25: 689). Having said this, it will be clear in the following that Kant sees *some* of these caricatured elements as having a good basis in Rousseau's own intellectual character.

been the hiding of true virtue beneath a layer of insincere behavior that has become impenetrable even to our own introspection:

What a train of vices must attend upon such uncertainty. No more sincere friendships; no more real esteem; no more well-founded trust. Suspicions, offenses, fears, coolness, reserve, hatred, betrayal, will constantly hide beneath this even and deceitful veil of politeness, beneath this so much vaunted urbanity which we owe to the enlightenment of our century. (Rousseau 1750/2008, 8, OC III, 8–9)²⁸

Rousseau's account of the role of the arts and sciences — philosophy included — is perhaps even more negative, in that they are in large part the cause of the corruption of morals. For Rousseau, "our souls have become corrupted in proportion as our Sciences and our Arts have advanced toward perfection" (Rousseau 1750/2008, 9 OC III, 9). The role of the sciences and arts is nothing more than to serve as the handmaiden of government, to keep the populace docile and distracted with trivialities so that they may be rendered more subservient. While government and society introduces chains that render human beings slaves, "the Sciences, Letters and Arts, less despotic and perhaps more powerful, spread garlands of flowers over the iron chains with which they are laden" (Rousseau 1750/2008, 6, OC III, 7). The goal of philosophy is to make slaves "love their slavery, and fashion them into what is called civilized Peoples" (*ibid.*).

The outputs of an enlightened cultivated reason are themselves *luxuries*, Rousseau insists (Rousseau 1750/2008, 18, OC III, 19). The introduction of luxury into modern society is itself the cultivation of

new desires demanding satisfaction.²⁹ However, the proliferation of desires cannot keep up with our capacity to satisfy them. As he puts it in *Émile*, "[s]ociety has made man weaker ... in making his strength insufficient for him" and this is just because "his desires are multiplied along with this weakness" (Rousseau 1762/1979, 84). The cultivation of reason often frustrates its own attempt to secure happiness. More importantly, though, they distract from genuine value, since "[m]inds debased by a host of futile cases cannot possibly ever rise to anything great" (Rousseau 1750/2008, 19, OC III, 20).

Rousseau sharply presents philosophers as prime offenders in the business of using rational capacities for the purposes of undermining morality — they "go off in all directions, armed with their deadly paradoxes; undermining the foundations of faith, and annihilating virtue" (Rousseau 1750/2008, 17, OC III, 19). He presents philosophers as sophists primarily motivated by the securing of a share of the marketplace of ideas, a "troop of charlatans, each hawking from its own stand on a public square" (Rousseau 1750/2008, 25, OC III, 27). Rousseau also mocks what he takes to be the patent absurdity of their views, such as "that there are neither virtues nor vices, and that moral good and evil are chimeras" (*ibid.*). He imagines descendants reading the works of Hobbes and Spinoza and declaring:

Almighty God, you who hold all souls in your hands, deliver us from the enlightenment and deadly arts of our forefathers, give us back ignorance, innocence and poverty, the only treasures that can make us happy and that are precious in your sight. (Rousseau 1750/2008, 26, OC III, 28)³⁰

The impact of Rousseau's writings upon the second half of 18th-century intellectual culture is difficult to overstate. The positive significance of

28. Perhaps echoing La Rochefoucauld, Rousseau claims that that no one any "longer dares to appear what one is" (*ibid.*, 8). La Rochefoucauld's scepticism can be thought of as more than the supposition that some apparent virtues are in fact disguised vices, but as also including the claim that we do not know ourselves (La Rochefoucauld 2008, V: 119). Kant himself seems to have thoroughly integrated this particular pessimism regarding introspective access to our motives in the *Groundwork*, e.g. (4: 407).

29. See (Velkley 2013, 93); also (Garrard 2003; Hulliung 1994; Mendham 2010).

30. Rousseau concludes with an opposition between Athenians and Spartans, the former who knew how to 'speak well' but the latter who knew how to 'act well' (Rousseau 1750/2008, 28, OC III, 30).

Rousseau for Kant himself during the 1760s was enormous. Yet given what we know about the basic orientation of the later Critical project, it is obvious that Kant came — probably sometime in the late 1760s — to reject the Rousseauian claims regarding the devaluing of rationality and philosophy, and instead sought to return them both to the center of human life. To give a single but important example of the pervasiveness of those claims in that period, Kant's own student Herder uncritically represented the same Rousseauian theses (and even in somewhat derivative prose) in the essay, *How Philosophy Can Become More Universal and Useful for the Benefit of the People*:

The highest degree of philosophical ability cannot at all coexist with the highest level of the healthy understanding; and so the dissemination of the former becomes harmful for the people. As soon as our soul transcends the bounds of need, it is insatiable in the desire for excess, and if philosophy determines nothing essential in what is necessary, then it is among those sciences which never allow an end of curiosity. (Herder 1765/2010, 11)

In short, O philosopher, go to the country and learn the way of the farmers, refine this picture into an ideal, and overthrow the unphilosophical manner of living, overthrow the idol which shows you philosophy as corruption of the world, but not through philosophy. (Herder 1765/2010, 23)

Herder's rhetoric here is a self-avowed endorsement of the characterization of philosophical theorizing as a luxury that goes beyond worldly need and recommends the rejection of this "luxury" through non-philosophical engagement with ordinary life. The challenge Kant faced then concerned how to accommodate Rousseau's premise regarding the reliability of unreflective moral consciousness without it

entailing Rousseau's conclusion regarding the rejection of philosophical reflection as relevant to ordinary moral life.

4. The Threat of Misology

Rousseau's influence is evident early in the text of the *Groundwork* with regard to the threat of misology. The account turns on an unquestioned acceptance of Rousseau's claim that the cultivation of reason in modern society has effected a production of countless new desires. The consequence of this modern phenomenon is the impossibility of satisfying them all and entails that the securing of happiness is perpetually deferred:

In actual fact, we do find that the more a cultivated reason engages with the purpose of enjoying life and with happiness, so much the further does a human being stray from true contentment; and from this there arises in many, and indeed in those who are most experienced in its use, if only they are sincere enough to admit it, a certain degree of *misology*, i.e. hatred of reason, since after calculating all the advantages they derive — I do not say from the invention of all the arts of common luxury, but even from the sciences (which in the end also appear to them to be a luxury of the understanding) — they still find that they have in fact just brought more hardship upon their shoulders than they have gained in happiness, and that because of this they eventually envy, rather than disdain, the more common run of people, who are closer to the guidance of mere natural instinct, and who do not allow their reason much influence on their behavior. (4: 395–6)³¹

31. As Horn characterizes it:

When these people retrospectively reflect on the gains and losses of having cultivated their intellectual abilities within their biographies, they typically come to the conclusion that this development didn't lead them to a larger amount of happiness, but to an increase of hardship.

The Rousseauian themes and language are echoed in articulation of the misological condition. Firstly, there is the identification of the cultivation of reason as ultimately productive of problems with regard to contentment. Secondly, there is the explicit identification of arts and sciences as both being conceptualizable as luxuries. Kant thus directly engages with Rousseau (and Herder's) co-opting of the luxury debate of the early 18th century and presentation of philosophy as another luxury commodity.³² Thirdly, there is the Rousseauian resolution of a return to "natural instinct" still more observable among "the common run of people" and the entailed dethronement of reason in guiding moral behavior.

What are we to make of apparently misological judgments, ones that downgrade the efficacy of reason? While acknowledging the reality of the phenomenon that Rousseau identifies, Kant considers it an insufficient justification for the general dissatisfaction with reason itself. Kant's diagnosis is that the misological tendency has led agents to reason as follows: (i) nature allocates the proper faculty to the proper end of human beings;³³ (ii) the securing of happiness is the proper end of human beings; (iii) our rational capacities are not reliably connected to the securing of happiness; therefore, (iv) the proper end of

So they feel even envy for ordinary people and their non-intellectual, purely sensual way of life. (Horn 2006, 46)

This point is also noted also by (Shell 2009, 59).

32. (Louden 2010) puts weight on relevance of the *Second Discourse*, but it is especially the claims in the *First Discourse* that Kant is alluding to with references to the arts and sciences and the luxuries of the understanding. Rousseau is generalizing a common theme of the luxury debate of the 18th century so as to generate a counter-enlightenment critique of modern society (see (Berg and Eger 2002; Jennings 2007)). The philosophical uses of the luxury debate have been particularly well-explored by (Hont 2006, 2010).

33. Kant would have been recently familiar with the view from Herder's *Ideas for the Philosophy of the History of Humanity*, which he read and reviewed during the same period in which the *Groundwork* was published. There, Herder analyzed the proper function of animal sensation to the greater good of the species in detail, leading him to declare: "hail, then, overpowering instinct, infallible guide!" (Herder 1803, 108). For discussion of Kant's break with Herder see (Ameriks 2012c).

human beings must be conducted through the use of the non-rational faculties that nature has bestowed.³⁴

The thought that human nature might not be essentially marked out in its moral dimension by virtue of its exercise of rational capacities would have been a familiar one by the time of the composition of the *Groundwork*.³⁵ Kant's claim here is an attempted diagnosis of the reasons and motives for ending up with such a picture. He claims that such arguments presuppose a thesis regarding "the wisdom of nature" (4: 396). One might only downgrade reason within the human framework if one thought that there was some proper division of labour by nature with regard to which faculty one ought to attend in pursuing one's moral aims. It is for this reason that Kant suggests that the misological tendency is "by no means sullen, or ungrateful to the kindness of the government of the world" (4: 396). The misologist's argument depends upon the wisdom of nature thesis in order to generate the worry that reason seems to fail to perform the assumed task of moral governance.

The argument here is often dismissed as a piece of lame teleological reasoning entirely unwelcome within the Critical philosophy.³⁶ However, it is important to note the dialectical position of the argument in the text. Kant is clearly arguing here on his opponent's own terms. The misologist correctly notes that when reason moves us to action, it can do so in ways that are not conducive to our happiness. Kant's response

34. For analysis see (Allison 2011, 80–6; Schönecker and Wood 2015, 47–50; Sedgwick 2008, 53–5; Timmermann 2007a, 22–4). Allison calls this portion of the *Groundwork* Kant's "teleological interlude" (Allison 2011, 81), and suggests that its presence is largely for the purpose of a response to Christian Garve, though he acknowledges that on occasion Kant is "[e]choing Rousseau" (Allison 2011, 83). However, Allison does not draw the connection between the analysis presented here and their oppositions to specific claims of the *First Discourse*.

35. Apart from Hume, Kant would have identified something like this position with Montaigne and Mandeville — see *Practical Reason* 5: 40.

36. E.g. (Allison 2011; Paton 1946; Timmermann 2007a; Wolff 1973). However, for an interesting discussion — one that notes the importance of both Rousseau and the *Phaedo* (though different aspects of the latter are focused upon that are here) — see (Horn 2006).

agrees with this central premise, since by doing so, it reveals that his opponents also concede a crucial claim, namely that reason is capable of being practically efficacious. It is crucial for Kant's larger goals — only explored in the Second and Third Sections of the *Groundwork* — that *the very idea* that our behavior can be motivated by purely rational means is granted. Kant notes that the argument depends on the claim that "reason as a practical faculty, i.e., as one that is meant to influence the *will*, has yet been imparted to us" (4: 396 — emphasis in original).

The rhetorical aim is to show that the possibility of some form of rational agency is accepted on all sides. The complaint of the misologist is not that we are afflicted with an idea of rational motivation that is itself a phantasm. Instead, the argument assumes the possibility of the motivational power of our rational capacities and questions whether, given the fact that reason and desire can *both* influence the will — and moreover can point the will in opposed directions — the better explanation of these facts is that the cognitive development of human beings in society has taken a wrong turn at some point. Given that the aim of human existence is happiness, the opponent reasons, and given that instinct's motivational efficacy is reliably connected to the securing of happiness, the tendency to accept the apparent authority of our rational capacity and to let it rule our practical lives should be resisted. In this way, the lack of integration between reason and instinctual desires is presented as a late-stage cultural crisis of the self clearly recommending a kind of primitivism.

The problem with this argument (as Kant has already argued at 4: 393) is with the identification of the characteristic good of human existence with the securing of happiness. Only if one makes this identification might one be subsequently troubled by the fact that our rational capacities often hinder as much as they help the securing of happiness. If the securing of happiness is the *raison d'être* of a human being, then the practical deliberations that form the core part of human beings' existential conditions are recast in a deeply problematized manner. If happiness is the end of human beings, then they are profoundly self-deceived with regard to their disposition to give default authoritative

weight to the ordinances of the rational side of the wills whenever they arise.³⁷

5. Socrates, Reason, and Misanthropy

One striking but under-discussed fact about Kant's analysis here is his use of the word 'misology'.³⁸ It is well known that the term stems from Plato's *Phaedo*, yet the significance of this origin is rarely touched upon. One reason for this might be the assumption that Kant was continuing the vogue for the dialogue inaugurated by the enormous success of Mendelssohn's *Phädon* in 1767, a part-translation-of, part-excursus-upon the *Phaedo*. However, although the relevant passage is translated by Mendelssohn, there he describes the risk of becoming a "hater of reason" [*Vernunftthasser*] — Mendelssohn does not use the term 'Misologie', as Kant does.³⁹ Kant makes a (for him quite rare) return to Plato's text as a source for his point here. I would argue that Kant turns to Plato at this point in the text specifically again as a response to Rousseau.

In the *Phaedo*, Socrates takes a pause in the argument to offer a warning against drawing too dismal a conclusion:

'But first let's take care that a certain fate doesn't befall us.'
'What's that?' I asked.

'The fate of becoming "misologists", just as some become misanthropists; because there's no greater evil that could befall anyone than this — the hating of arguments. Misology and misanthropy both arise from the same source. Misanthropy develops when, without skill, one puts complete trust in somebody, thinking the man absolutely true and sound and reliable, and then a little

37. Of course Kant, too, presents an existential account of human beings whose proper condition is one of internal conflict. His aim, I would claim, is not to replace this picture, but to properly cast the protagonists and antagonists within that picture.

38. Kant sometimes uses the term frequently to describe philosophical schools of a broadly naturalistic or empiricist disposition, e.g. *Logic*, 24: 36.

39. (Mendelssohn 1767/2006, 109–10).

later finds him bad and unreliable; and then this happens again a little later with another person; and when it happens to someone often, especially at the hands of those he'd regard as his nearest and dearest friends, he ends up, after repeated hard knocks, hating everyone, thinking there's no soundness whatever in anyone at all.' (*Phaedo*, 89c–d)⁴⁰

There are several important points to note here. Firstly, Plato links the hatred of reason with the hatred of mankind itself. Secondly, and more importantly though, is the fact that development of both misology and misanthropy is the result of a kind of fallacy. With regard to misology, one initially has some bullish trust and confidence in reason, only for one's expectations to be dashed when one finds that reason produces aporiae, paradoxes, and antinomies. The fallacy occurs when one infers from this result that reason itself is generally unreliable. The proper response, Socrates suggests, is to re-examine oneself and one's own handling of one's rational capacities. It is the mistake, then, of confusing an operator error with a system error. Misanthropy operates in the same way, whereby one encounters disappointment with regard to one's high expectations of other human beings, and rather than revise one's expectations, one instead infers that the species as a whole lacks value.⁴¹

Kant would have been familiar with another different account of the origin of misanthropy from Cicero's *Tusculan Disputations*.⁴² Here, misanthropy is paired instead with misogyny:

It is thought moreover that fear is the origin of their opposites like *hatred of women*, as for instance in the *Μισόγυνος* of Atilius, like the *hatred of all mankind* felt we are told

40. For discussion, see (Jacquette 2014; Miller 2015; Scott 2006, 73; Woolf 2008).

41. See also *Logic* 24: 204, 24: 800, 24: 74; *Metaphysics*, 28: 535.

42. For some of the influences of Kant's reading of Cicero see (Doyle and Torralba 2016; Schneewind 2009). For a single example, Kant refers to an anecdote from the *Tusculan Disputations* in the *Critique of Practical Reason* at 5: 60.

by Timon who is termed *μισάνθρωπος*, and like *inhospitality*: and all these sicknesses of the soul originate in a certain fear of the things they avoid and hate. (Cicero, *Tusculan Disputations* IV.xi, 35)

For Cicero, the origin of misanthropy and misogyny is more straightforwardly generated from a negative attitude towards human beings and women, respectively. When discussing misology, it is striking that Kant groups it together both with misanthropy and misogyny.⁴³ However, he provides an analysis of their generation that clearly derives from Plato rather than Cicero. Writing to Herz about their mutual acquaintance Kraus, Kant writes:

... A certain misology that you, as I, detected — and regretted in Mr. Kraus derives, as does much misanthropy, from this: that in the first instance one loves philosophy, in the second, people, but one finds both ungrateful, partly because one expected too much of them, partly because one is too impatient in awaiting the reward for one's efforts from the two. I know this sullen mood also (Letter to Marcus Herz, February 4, 1779, quoted in (Kuehn 2001, 210))

This attitude (one Kant confesses to have previously maintained himself) whereby the hatred of the thing stems from one's previous love of that very same thing and which generated unrealistic and unrealized high expectations for that thing, is repeated in many of Kant's writings. In the lectures on anthropology, he mentions misology, misanthropy, and misogyny together:

If reason just cannot fulfill knowledge, if it cannot satisfy the individual in this, if it deserts him in this, so that the

43. (Shell and Velkley 2017, 204–05) and (Ameriks 2017) are the only commentators I am aware of who note Kant's connection of misanthropy, misology, and misogyny and their relation to Kant's engagement with Rousseau, though they don't explore its classical origins as I do here.

individual cannot foresee the goal and end of all things, then the individual resorts to simplemindedness, and renounces reason altogether, just as someone becomes a misanthrope due to the sensation of virtue, not because he despises people, but because he does not find them to be how he wants them to be. ... Thus one also does not become a misologist out of hatred for reason, indeed one values it, but because it does one a disservice, one thus renounces it. ... Misogyny, or hatred of women, occurs in the same way. It also arises from an ill humor, not because one despises them, but because one does not find in them what one believes, thus from an entirely too great a demand for their perfections. (*Anthropology*, 25: 552–3).

Kant takes pains to point out the benevolent origins of these attitudes, chastising it only for its over-exaggerated enthusiasm “since if such an ideal is not attained, then such an enthusiasm produces misanthropic individuals” (*Anthropology*, 25: 531). Neither does Kant hesitate in identifying a clear case:

Such enthusiasts are not malicious people, but they are touched with principles of benevolence toward the entire human race, and since they cannot find such, they become misanthropes, for example, Rousseau (*Anthropology*, 25: 530)⁴⁴

Kant’s own invocations of Socrates are plausibly understood as made in deliberate opposition to Rousseau’s own invocation. In the *First Discourse*, Rousseau holds up Socrates as the example of the *anti-philosopher*, who challenged poets, artists, and orators in turn as sources

44. Further references to Rousseau’s misanthropy can be found in *Anthropology*, 25: 846, 25: 1302, 25: 1364. He also refers to Rousseau as a “subtle Diogenes” (*Anthropology*, 25: 724), an allusion that none of his readers could have thought positive. Voltaire, among many others, had levelled the same charge against Rousseau (Hullung 2001) — on Rousseau and misanthropy, see also (Evrigenis 2010).

of virtue, only to find them all lacking.⁴⁵ Socrates “would continue to despise our vain sciences” were he to see them today, Rousseau maintains (Rousseau 1750/2008, 13 OC III, 30). In refusing to state a definition of virtue that might allow for some explicit decision procedure for moral guidance, Rousseau sees Socrates as manifesting wisdom in leaving us a form of teaching only in “the example and memory of his virtue” (*ibid.*).

Kant acknowledges that *in theory* one can proceed without “need of science and philosophy to know what one has to do in order to be honest and good, indeed even to be wise and virtuous” (4: 404). He grants that something like an expert interrogation of the fine grain of moral scenarios without explicit appeal to systematically elaborated principles *could* keep one on the path of virtue. Common reason could manage this if one could, “as Socrates did, make it aware of its own principle” (4: 404). Kant’s point though is that this is possible only for someone with Socrates’ gift for remaining both purely rational and focused on the concrete details of the scenario. For most of us — and most importantly, for most of the ordinary people Rousseau and Herder are asking us to emulate — this procedure will be far less reliable. In this way, Kant is cleverly claiming that ironically there is an implicit elitism in Rousseau’s proposal, as only someone with philosophical understanding of Socrates’ method could benefit from his recommended procedure.

Kant also acknowledges Rousseau’s challenge more explicitly when he asks whether it might not be “more advisable, in moral things, to leave it with the judgment of common reason” (4: 405). He acknowledges that reason itself can lead one away from virtue when it “becomes subtle” and on occasion engages in “legalistic quibbles with its own conscience” (*ibid.*). Does this entail that we ought to abandon philosophy and, as Rousseau pleads, be returned to “ignorance, innocence and poverty”? Again, Kant is clear in opposing Rousseau’s linking of innocence and natural “wisdom” [*Weisheit*] on the one hand

45. For a detailed account of Rousseau’s use of the image of Socrates, see (Orwin 1998).

with the lack of learning or “knowledge” [*Wissen*] on the other.⁴⁶ The idea that we might be free from moral corruption if only we were free from cultivated reason is presented as a naïve myth. Kant’s implication is that Rousseau’s aspiration to innocence, while admirable, is premised on a false picture of human beings as incorruptible in their “natural” state:

Innocence is a glorious thing, but then again it is very sad that it is so hard to preserve and so easily seduced. Because of this even wisdom — which probably consists more in behavior than in knowledge elsewhere — yet needs science too, not in order to learn from it, but to obtain access and durability for its prescription. (4: 404–5)

Kant is thereby turning Socrates against Rousseau by claiming that the latter has failed to heed the former’s warning about the misological fallacy. Rousseau has laid a charge against reason, Kant is claiming, as a result of Rousseau’s own failure to reason correctly regarding the nature of human innocence and corruption. This distinction between the need for first-order instructions and a more reflective need for “access and durability” is then crucial for Kant’s anti-Rousseauian defence of the need for philosophy “in moral things”.

6. The Natural Dialectic

While Kant accepts Rousseau’s claim that philosophy is not required for first-order moral guidance, he denies that endorsement of this claim entails that philosophy is thereby not required in order for the cultivation of one’s moral life more generally.⁴⁷ Kant’s claim is that a

46. Kant connected the idea of moral philosophy as inculcating a “wise innocence” as early as 1765 in the announcement of his lectures for 1765–6 (Kant 1992, 2: 311–2). Kant warns about the frailty of innocence in (Kant 2008, 29: 604).

47. One can also distinguish different forms of corruption here, from the tendency to think that moral demands are not authoritative to the tendency to think that they are authoritative but are nevertheless selectively applicable and capable of gerrymandering.

natural dialectic can arise that can undermine or corrupt one’s own moral integrity, and which demands philosophical inquiry as a response. Given the preceding analysis, one can see that the natural dialectic is properly thought of as a manifestation of misology, since it is an analysis of the ways in which we come to distrust the rational side of our moral consciousness.⁴⁸ The process is described in a characteristically compressed passage, one that requires careful reconstruction:

The human being feels within himself a powerful counterweight to all the commands of duty — which reason represents to him as so worthy of the highest respect — in his needs and inclinations, the entire satisfaction of which he sums up under the name of happiness. Now reason issues its prescriptions unrelentingly, yet without promising anything to the inclinations, and hence, as it were, with reproach and disrespect for those claims, which are so vehement and yet seem so reasonable (and will not be eliminated by any command). But from this there arises a *natural dialectic*, i.e. a propensity to rationalize against those strict laws of duty, and to cast doubt on their validity, or at least their purity and strictness and, where possible, to make them better suited to our wishes

48. Wood implies that the reason why a subject is inclined to moral self-deception is due to the natural drive to “assert their self-worth antagonistically in relation to others” (Wood 2008, 6). The idea is presumably that we naturally hold our ends to be more valuable than the ends of others, and since morality often demands that we value the ends of others over our own, we have an inevitable tension that can drive one to dissemble in the favour of self-interest. For this and other reasons, Wood sees Kant as fundamentally in accord with Rousseau with regard to the societal source of moral corruption (Wood 1999, 2008, 2010). There is no doubt that Kant does think that human beings have this natural self-aggrandizing tendency. However, it is notable that there is no indication that *this* claim regarding human nature appears to be doing any work in the natural dialectic passage. Rather, here Kant seems to say that there is just something about the phenomenology of moral demands that can lead to the natural dialectic. However, a fuller discussion of this matter, which ultimately concerns the relation between natural dialectic and radical evil, will have to be conducted elsewhere.

and inclinations, i.e. fundamentally to corrupt them and deprive them of their entire dignity, something that in the end even common practical reason cannot endorse. (4: 405)

There are several claims made within the brief passage, which is a perspicuous and imaginative reconstruction of the psychology of moral self-deception. The first thing to note is that Kant characterizes the situation as one where there are two sets of competing claims, one from reason and the other from inclination, and hence two voices within a single consciousness.⁴⁹ Kant's formulation of the problem is in terms of a self who is analyzing the pattern of one's own internal practical deliberations.

Secondly, Kant points out that the fact that reason's prescriptions are categorical in nature — as he has been claiming throughout the First Section — involves them having the peculiar character of denying inclination any reward that might also be satisfied by virtue of those prescriptions being followed. This after all was the exact character of categorical commands that was identified as the essential feature of the sense of duty: When reason prescribes a course of action, then it does so on the sole grounds that the *action simply is the right thing to do* — the prescriptions of reason appear to ordinary consciousness with an almost dogmatic character. They simply state that something is or is not to be done without ever stating what advantage accrues if that thing is done or not done.⁵⁰ It is also in the nature of such commands that they arise “unrelentingly”, i.e. with an unvarying constancy that pays no heed to the vagaries of context, to the ease or difficulty

49. Ware denies that this characterization of the sceptical threat is an “adversarial” one, on the grounds that “the skeptic most worth addressing lies within ourselves” (Ware 2014, 376). While it is certainly correct that the sceptical challenge is viewed by Kant as one that is part of our inner conflict, it is surely also the case that he presents the situation as one where two sides of our own nature compete as adversaries.

50. For a recent discussion of the “Motive of Duty Thesis”, see (Markovits 2010). In MacIntyre's memorable phrase, moral prescriptions “do not enjoin us hypothetically; they simply enjoin us” (MacIntyre 2007, 44).

with which those commands might be obeyed, or to the subtleties of personal circumstance. Kant's claims regarding the first-personal phenomenology of categorical claims *entails* that the subject is presented with commands without any grasp of what is gained by following them. The provision of such a gain would of course be to transform categorical imperatives into hypothetical ones (4: 414). This is one sense in which the dialectic is natural, since it is engendered by the fact that duty enjoins us to action categorically and *this* fact entails the absence of any second-order narrative as to why first-order commands do in fact enjoin us categorically.

Thirdly, Kant also notes several key features of the phenomenology of inclination. In the first place, our inclinations are “vehement” in their own particular way; they make an appeal on our consciousness that can seem just as natural — and can on occasion seem to convey a comparable urgency — to that of the commands of reason. Moreover, our inclinations “seem so reasonable”: our desires and inclinations, when they influence us, always do so under the guise of the good.⁵¹ When we desire something, it is constitutive of the experience that the realization of that thing is presented to consciousness as something that would be good for the agent. While there are morally good and bad states of the will, *that* one has the mere desire is not itself bad. It does not seem to the subject that the mere presence of a desire itself reflects a moral failing on their part. Yet the absence of any second-order narrative as to the source of moral commands entails that we lack any account of *why* our desires make recommendations under the guise of the good that yet are on occasion properly prohibited by a different internal voice.

Fourthly, as Kant points out, the fact that the satisfaction of some desires is on occasion judged inappropriate does not thereby eliminate that desire from my consciousness. Even if resisted, the voice of desire is not thereby silenced. One can still maintain the “goodness” of the desire even subsequent to the issuance of a moral command

51. See (Velleman 1992; Tenenbaum 2010).

that one ought not to satisfy that desire on that occasion. Furthermore, there are other contexts where the condemning voice of reason remains silent and one's pursuit of that very same desire is tacitly sanctioned. Yet reason again does not on any of these occasions issue any explanation regarding how to accommodate these facts within one's consciousness. Reason's voice can then seem to be heard only on occasions where there is some potential conflict to be arbitrated, and the method of arbitration is always flat prohibition without negotiation. It insists categorically *that* occasional prohibitions of acting on our desires are necessitated for moral propriety, but does not offer any explanation that might explain just *why* the unquestioning restraint of desire by reason might be the proper way for a human being to behave.

It is in these senses that reason seems to offer "reproach and disrespect" to the faculty of inclination itself. By only issuing forbidding commands and by failing to offer any explanation for how — or even why — one's faculty of inclination is to be integrated with the faculty of reason, *reason itself* creates the conditions for internal existential discord. These are the factors that account for the natural dialectic arising. Without reflection as to why this inner condition must be as it is, the subject can come to regard the voice of reason as nothing but an arbitrary authority figure. The lack of explanation as to its role within the broader narrative of one's existence can lead the subject to raise the possibility that its authoritative status is unearned and unwarranted. Given its unwillingness to explain its judgments to inclination, it can seem that reason proceeds to wield that power in policing our inclinations with a kind of contingency. The generation of this impression is again *natural* just because it arises as a result of our faculties of inclination and reason operating in their proper manner. It is a *dialectic* because when these two faculties operate just as they should they then create conditions whereby the claims of the one is challenged by the claims of the other.⁵²

52. It is notable then that, contrary to (Allison 2011, 143–45), there is a sense in which an antinomy arises such that Kant's referring to the situation as a dialectic is not "artificial". There are of course important differences between this

One might object that a different reading of the passage is available, whereby the process undergone by the subject is not one whereby the authority of moral demands is undermined.⁵³ On this account, the subject retains the initial belief that the rational demands of morality are authoritative in all cases, but comes to doubt whether her failure to regard them as such is really a culpable failure. One rationalizes that, even though one ought to follow the demands of morality in all cases, there are occasions where one fails to do so in ways that are genuinely non-blameworthy. In other words, she rationalizes that not all cases of moral *akrasia* are blameworthy. Such a reading does not reflect the characterization of the natural dialectic that Kant offers however. The subject engages in a process of discovering "legalistic quibbles" with regard to the demands of morality; she seeks to find interpretations of those demands that undermine their "purity and strictness" and "are better suited to our wishes and inclinations" (4: 405). The description of the dialectic is one whereby the laws *themselves* are reconceived as not applying strictly, i.e. as applying to us only in certain cases. This is of course different from the claim that the laws do apply strictly but that our failure to obey those laws is non-blameworthy.

It is important for Kant's account of the psychology of moral corruption that it captures the sense in which it is distinct from an explicit rejection of morality *per se*. The response that Kant is interested in though is not one whereby a subject considers whether the institution of morality is itself an illusion and that one perhaps ought to do what one pleases. The response that interests Kant is a kind of self-deception whereby the subject undermines morality while still paying lip service to it. The subject does not claim that morality is not binding on them, but instead reinterprets what it is for morality to be binding in a way that allows for its "strict" and "unrelenting" dictates to be avoided on occasion.

and the antinomies of the First *Critique*, not least that the latter are understood as problems regarding the production of equally valid yet opposing claims by the faculty of reason alone.

53. I am grateful to Yoon Choi for pressing me to clarify my reading in this regard.

This kind of psychological self-deception is repeated in Kant's well-known characterization of what occurs when a subject contemplates a maxim that in fact fails the categorical imperative procedure. There, he describes a very similar psychology, whereby one takes oneself to have quibbled with the strictness of the moral law in a way that allows one to make an exception for oneself, while at the same time maintaining that one is perfectly responsive to morality, since one *would* follow the moral law if only it applied in this case:

If we now attend to ourselves in every transgression of a duty, we find that we actually do not will that our maxim should become a universal law, since that is impossible for us, but that its opposite should rather generally remain a law; we just take the liberty of making an exception to it for ourselves, or (just for this once) to the advantage of our inclination. (4: 424)

7. Conscience and the Voice of Reason

The characterization of the natural dialectic I have offered throughout this paper has been in terms of an internal negotiation between two competing voices. This characterization is one Kant was familiar with, and the passage can be seen as a direct response to, and re-characterization of, Rousseau's account of the same struggle detailed by the Savoyard Vicar in *Émile*. One of the vicar's "articles of faith" concerns "what rules I ought to prescribe for myself in order to fulfil my destiny on earth" (Rousseau 1762/1979, 286).⁵⁴ The vicar assumes that God has granted him "conscience to love, reason to know and, liberty to choose" without extended or sophisticated reflection:

In continuing to follow my method, I do not draw these rules from the principle of a high philosophy, but find

54. Kant explicitly refers to the Savoyard Vicar for example in *Anthropology from a Pragmatic Point of View*, 7: 326–27. Much like the Savoyard Vicar, Kant concludes in the First *Critique* that the reality of God and our immortal souls are "articles of faith" (A830/B858).

them written in the depth of my heart. I have only to consult myself about what I want to do. (Rousseau 1762/1979, 286).⁵⁵

However, Rousseau goes on to characterize the rationalization against conscience:

The best of all casuists is the conscience; and it is only when one haggles with it that one has recourse to the subtleties of reasoning. The first of all cares is care for oneself. Nevertheless how many times does the inner voice tell us that, in doing our good at another's expense, we do wrong! We believe we are following the impulse of nature, but we are resisting it. In listening to what it says to our senses, we despise what it says to our hearts; the active being obeys, the passive being commands. Conscience is the voice of the soul; the passions are the voice of the body. Is it surprising that these two languages often are contradictory? And then which should be listened to? Too often reason deceives us. We have acquired only too much right to challenge it. But conscience never deceives; it is man's true guide. It is to the soul what instinct is to the body; he who follows conscience obeys nature and does not fear being led astray. (Rousseau 1762/1979, 286–87)

Rousseau thus sees the problem as occurring at the very first step, when a subject even deigns to "haggle" with the first-order prescriptions of conscience. Here though, the voice of conscience is put in contrast not just with the voice of the passions but also with the voice of reason. Reason's voice is too often recruited on behalf of the passions in order to disguise vices as virtues and to justify to oneself actions that would otherwise be grasped as going against conscience. The result is an inversion of the natural hierarchy of a subject's faculties. Throughout *Émile*, Rousseau articulates the natural superiority of

55. For the anti-philosophy theme, see (Rousseau 1762/1979, 268 ff.).

the “active” faculty of the human being in contrast to the passive faculties of sensation and the passions. The distinctness of human beings is defined in these terms, since “no material being is active by itself” (Rousseau 1762/1979, 280), but human beings have an active principle expressible in their free will.

The very goal of *Émile*’s education is to bring this disposition to fulfilment and to create an “active and thinking being” (Rousseau 1762/1979, 203). Nature has allowed this superiority of our free wills to be evident to us in the dictates of conscience, when the latter prescribes against the fulfilment of desire. When one rationalizes against the voice of conscience however, one inverts the natural order of the authority of one’s inner voices. In such a case, one takes desires to be authoritative just because they arise as seemingly natural “impulses”, and one rationalizes that one is following nature in following those impulses. In reality though, since the authority of the voice of conscience is in fact itself the proper endowment of nature, we are “resisting” nature when we follow our natural impulses. The structure of governance of the self is then perverted against nature, with the lower passive self issuing commands, and the higher active self obeying them.

Kant’s natural dialectic passage is clearly a reconceptualization of this one. His response is not to deny that such rationalization can take place, but only that such rationalization entails neither that philosophical reasoning should be forsaken nor that conscience should be trusted blindly and without reflection on its nature. On the contrary, it is blind obedience to the dictates of conscience itself without an explanation of the rational nature of its authority that generates the conditions in which the disposition to disobedience can flourish. Furthermore, the assumption that we have a reliable and common capacity for first-order moral prescriptions does not entail that we ought to privilege it over our rational capacities, because for Kant, reason itself can be *identified* as that reliable first-order common capacity. If this is granted, then the acceptance of the possibility of common moral cognition need not put it on a collision course with the authority of

rationality, and need not demand a characterization of rational faculties as a resource for self-deception. On the contrary, occasions of self-deception are properly characterized as occasions of *inclination*’s co-opting of reason’s resources against reason itself.

With the Rousseauian context in mind, we can grasp some answers to the questions originally raised by the natural dialectic passage. Firstly, there is a straightforward account of how Kant can maintain the necessity of philosophy in the face of the sufficiency of common moral cognition. Common moral cognition is sufficient to give reliable responses to moral scenarios when it attends to the voice of reason. The voice of reason though speaks categorically at the first-order level and so does not offer any explanation of why and how it is reliable. The very categorical character of our first-order moral responses — the feature of rational prescriptions that they merely bluntly state *that* they are authoritative — entails that we lack a sense at the first-order level as to *why* they are authoritative first-order moral responses. This absence creates space for doubt. Philosophical inquiry is necessitated not to provide us with extra first-order guidance, but rather with second-order understanding *that* common moral cognition is sufficient at the first-order level.

Secondly, there is an equally clear sense of how second-order reflection is now necessitated. The rationalizing that inclination engages in is a piece of *theorizing*, i.e. it must present to consciousness the thought that moral commands are not strictly universal in order for it to have its corrupting efficacy upon the will. More importantly, Kant seems to imply that this state of corruption is internally unstable. The process of the rationalizing out of moral obligations is “something that in the end even common practical reason cannot endorse” (4: 404). Presumably, the thought is that with regular rationalizing out of particular obligations on particular occasions, one gradually loses a sense of the authoritativeness of moral commands as a *general* feature of them. One loses one’s grip on the very distinction between moral and non-moral commands. This, however, is antithetical to the original aims of the self-deceiving subject, which was not to reject morality *per se*

but rather to find some wriggle room for preferred courses of action while maintaining an image of oneself as committed to morality. It is for this reason that Kant says that “*common human reason* is impelled to leave its sphere not by some need of speculation (which never comes over it as long as it is content to be mere sound reason), but rather on practical grounds” (4: 405 — emphasis in original). In this manner, Kant turns Rousseau’s premise on its head. The unreflective character of common moral cognition is not a basis for the irrelevance of philosophical reflection; on the contrary, it is its very unreflective character that engenders the moral aporia and the subsequent drive to reflection.

Thirdly, the analysis of the natural dialectic points the way as to how philosophical inquiry could be *sufficient* to alleviate the corrupted state of the subject and ward off the threat of future corruption, thereby securing a more reliable “access and durability” for the first-order commands commonly understood. The threat comes from the lack of a narrative for why it could be that our immediate and unreflective moral responses might nevertheless be rational responses; how it could be that our rational responses really do provide an objective criterion for moral behavior; how it could be that those responses can on occasion speak against recommendations of inclinations; how it could be that those recommendations of inclination can be understood as aspects of one’s own self; how it could be that the voice of reason is nevertheless entitled to speak authoritatively since it speaks as part of the “higher” or more authentic part of the self; and ultimately, how it could be that reason could have the mysterious practical efficacy that it appears to have.⁵⁶

It is the provision of a narrative containing precisely these elements of the natural dialectic that follows in the remainder of the *Groundwork*. The immediate response of respect for the law is shown to be expressible in the categorical imperative test (4: 421); the explanation of our

56. My aim of course has been merely to ascertain how Kant thought such philosophical reflection might suffice to alleviate the threat of moral corruption. I have left untouched the question as to whether it is plausible that it in fact does.

moral responses is explained by our existence as a bifurcated self of which the rational part, located in the intelligible world, is viewed as the true self (4: 457); ultimately the metaphysics of transcendental idealism is supposed to provide a model both for how pure practical reason is possible and yet in some sense still mysterious (4: 458). Were the readers of the *Groundwork* to find themselves compelled by its arguments, they would see not only the relevance of philosophical inquiry to practical life; they would have come to see the possibility that their initial immediate moral responses, far from being a distraction from one’s higher rational nature, are in fact an expression of it. They would also have been provided with a vindication of philosophy as a necessity rather than a luxury of practical life. Finally, they would have been presented with a vindication of reason itself, and a basis to resist the threat raised in the First *Critique*, that of our *indifference* to the recommendations of reason in the determination of our lives (*Ax*).⁵⁷ This risk is the one Kant saw as the greatest threat to the Enlightenment project and is the one that the *Groundwork* is directed towards averting.⁵⁸

Bibliography

- Alberg, Jeremiah. 2015. “Metaphysics as Kant’s Coquette: Rousseau’s Influence on *Dreams of a Spirit-Seer*.” *Kantian Review* 20 (3): 347–71.
- Allison, Henry E. 2011. *Kant’s Groundwork for the Metaphysics of Morals: A Commentary*. Oxford: Oxford University Press.
- Ameriks, Karl. 2012a. “Kant, Human Nature, and History after Rousseau.” In *Kant’s Observations and Remarks: A Critical Guide*, edited by

57. See also *Anthropology*, 28: 540.

58. This paper has benefitted crucially from conversations with Angela Breitenbach, Yoon Choi, Alix Cohen, Sacha Golob, James Harris, Wayne Martin, Sasha Mudd, Becky Vincent, and Eric Watkins. It was developed in a series of talks given at the Institute of Philosophy in London, the University of Essex, the University of St. Andrews, the Kant Congress in Vienna, and Keele University. I am grateful to the audiences on each of these occasions for their feedback. I am grateful also to two anonymous referees for their very helpful comments.

- Susan Meld Shell and Richard Velkley, 247–65. Cambridge: Cambridge University Press.
- . 2012b. *Kant's Elliptical Path*. Oxford: Oxford University Press.
- . 2012c. "Kant's Fateful Reviews of Herder's Ideas." In *Kant's Elliptical Path*, 221–237. Oxford: Oxford University Press.
- . 2017. "Once Again: The End of All Things." In *Kant on Persons and Agency*, edited by Eric Watkins, 213–30. Cambridge: Cambridge University Press.
- Armstrong Kelly, George. 2001. "A General Overview." In *The Cambridge Companion to Rousseau*, edited by Patrick Riley, 8–56. Cambridge Companions to Philosophy. Cambridge: Cambridge University Press.
- Bayle, Pierre. 1991. *Historical and Critical Dictionary: Selections*. Indianapolis, IN: Hackett.
- Beiser, Frederick. 2017. "The Enlightenment and Idealism." In *The Cambridge Companion to German Idealism*, edited by Karl Ameriks, 2nd ed., 21–42. Cambridge Companions to Philosophy. Cambridge: Cambridge University Press.
- Berg, Maxine, and Elizabeth Eger. 2002. *Luxury in the Eighteenth Century: Debates, Desires and Delectable Goods*. Basingstoke: Palgrave Macmillan.
- Brink, David. 1997. "Kantian Rationalism: Inescapability, Authority, and Supremacy." In *Ethics and Practical Reason*, edited by Garrett Cullity and Berys Gaut, 255–91. Oxford: Oxford University Press.
- Callanan, John J. 2019. "Methodological Conservatism in Kant and Strawson." *British Journal for the History of Philosophy* 27 (2): 422–42.
- Cassirer, Ernst. 1963. *Rousseau, Kant, Goethe: Two Essays*. Translated by James Gutmann, Paul Oskar Kristeller, and John Herman Randall Jr. New York: Harper & Row.
- . 1983. *Kant's Life and Thought*. Translated by James Haden. New Haven: Yale University Press.
- Cicero, Marcus Tullius. 1927. *Tusculan Disputations*. Translated by J. E. King. Loeb Classical Library. Cambridge, MA: Harvard University Press.
- Copp, David. 1991. "Moral Skepticism." *Philosophical Studies* 62 (3): 203–33.
- Dent, Nicholas. 2006. *Rousseau*. London: Routledge.
- Doyle, Daniel, and Jose Torralba. 2016. "Kant and Stoic Ethics." In *Routledge Handbook of the Stoic Tradition*, edited by John Sellars, 270–83. London and New York: Routledge.
- Evrigenis, Ioannis D. 2010. "Freeing Man from Sin: Rousseau on the Natural Condition of Mankind." In *Rousseau and Freedom*, edited by Christie McDonald and Stanley Hoffmann, 9–23. Cambridge: Cambridge University Press.
- Garrard, Graeme. 2003. *Rousseau's Counter-Enlightenment: A Republican Critique of the Philosophes*. Albany, NY: State University of New York Press.
- Geiger, Ido. 2010. "What Is the Use of the Universal Law Formula of the Categorical Imperative?" *British Journal for the History of Philosophy* 18 (2): 271–95.
- Grenberg, Jeanine. 2013. *Kant's Defense of Common Moral Experience: A Phenomenological Account*. Cambridge: Cambridge University Press.
- Guyer, Paul. 2008. *Knowledge, Reason, and Taste: Kant's Response to Hume*. Princeton: Princeton University Press.
- Harman, Gilbert. 1977. *The Nature of Morality: An Introduction to Ethics*. Oxford: Oxford University Press.
- Henrich, Dieter. 1992. *Aesthetic Judgment and the Moral Image of the World: Studies in Kant*. Palo Alto, CA: Stanford University Press.
- Herder, J. G. 1803. *Outlines of a Philosophy of the History of Man*. Translated by T. O. Churchill. London: J. Johnson.
- Herder, Johann Gottfried von. 2010. *Philosophical Writings*. Edited by Michael N. Forster. Cambridge: Cambridge University Press.
- Hill Jr, Thomas E. 1985. "Kant's Argument for the Rationality of Moral Conduct." *Pacific Philosophical Quarterly* 66 (1–2): 3–23.
- Hohenegger, Hansmichael. 2012. "From Rousseau to Kant: A Case of Translatio Iudicii." In *Translatio Studiorum: Ancient, Medieval and Modern Bearers of Intellectual History*, edited by Marco Sgarbi, 209–24. Leiden: Brill.

- Hont, Istvan. 2006. "The Early Enlightenment Debate on Commerce and Luxury." In *The Cambridge History of Eighteenth-Century Political Thought*, edited by Mark Goldie and Robert Wokler, 377–418. Cambridge: Cambridge University Press.
- . 2010. *Jealousy of Trade: International Competition and the Nation State in Historical Perspective*. Cambridge, MA: Harvard University Press.
- Horn, Christoph. 2006. "Kant on Ends in Nature and in Human Agency: The Teleological Argument (GMS I, 394–396)" In *Groundwork for the Metaphysics of Morals*, edited by Christoph Horn and Dieter Schönecker, 45–71. Berlin: Walter De Gruyter.
- Hullung, Mark. 1994. *The Autocritique of Enlightenment: Rousseau and the Philosophes*. Cambridge, MA: Harvard University Press.
- . 2001. "Rousseau, Voltaire, and the Revenge of Pascal." In *The Cambridge Companion to Rousseau*, edited by Patrick Riley, 57–77. Cambridge Companions to Philosophy. Cambridge: Cambridge University Press.
- Jacquette, Dale. 2014. "Socrates on the Moral Mischief of Misology." *Argumentation* 28 (1): 1–17.
- Jennings, Jeremy. 2007. "The Debate about Luxury in Eighteenth- and Nineteenth-Century French Political Thought." *Journal of the History of Ideas* 68 (1): 79–105.
- Kant, Immanuel. 1992. *Theoretical Philosophy, 1755–1770*, edited by David Walford and Ralf Meerbote. Cambridge: Cambridge University Press.
- . 2005. *Notes and Fragments*. Edited by Paul Guyer. Translated by Curtis Bowman and Frederick Rauscher. Cambridge: Cambridge University Press.
- . 2008. *Lectures on Ethics*, edited by Peter Heath and J. B. Schneewind. Translated by Peter Heath. Cambridge: Cambridge University Press.
- . 2011. *Groundwork of the Metaphysics of Morals*. Edited by Jens Timmermann. Translated by Jens Timmermann and Mary Gregor. Cambridge: Cambridge University Press.
- . 1902– *Kants Gesammelte Schriften*, ed. Königlich Preussische Akademie der Wissenschaften, vols. 1–29. Berlin: de Gruyter.
- Kuehn, Manfred. 2001. *Kant: A Biography*. Cambridge: Cambridge University Press.
- La Rochefoucauld, François. 2008. *Collected Maxims and Other Reflections*. Translated by E. H. Blackmore, Francine Giguère, and A. M. Blackmore. Oxford: Oxford University Press.
- Louden, Robert B. 2010. "Evil Everywhere: The Ordinariness of Kantian Radical Evil." In *Kant's Anatomy of Evil*, edited by Sharon Anderson-Gold and Pablo Muchnik, 93–115. Cambridge: Cambridge University Press.
- MacIntyre, Alasdair. 2007. *After Virtue: A Study in Moral Theory*, 3rd ed. Notre Dame, IN: Notre Dame University Press.
- Markovits, Julia. 2010. "Acting for the Right Reasons." *Philosophical Review* 119 (2): 201–42.
- Mendelssohn, Moses. 2006. *Phädon, Or, On the Immortality of the Soul*. Translated by Patricia Noble. New York: Peter Lang.
- Mendham, Matthew D. 2010. "Enlightened Gentleness as Soft Indifference: Rousseau's Critique of Cultural Modernization." *History of Political Thought* 31 (4): 605–37.
- Miller, Thomas. 2015. "Socrates' Warning Against Misology." *Phronesis* 60 (2): 145–79.
- Muchnik, Pablo. 2010. "An Alternative Proof of the Universal Propensity to Evil." In *Kant's Anatomy of Evil*, edited by Sharon Anderson-Gold and Pablo Muchnik, 116–43. Cambridge: Cambridge University Press.
- Neuhouser, Frederick. 2008. *Rousseau's Theodicy of Self-Love: Evil, Rationality, and the Drive for Recognition*. Oxford: Oxford University Press.
- Orwin, Clifford. 1998. "Rousseau's Socratism." *The Journal of Politics* 60 (1): 174–87.
- Paton, H. J. 1946. *The Categorical Imperative: A Study in Kant's Moral Philosophy*, vol. 58. London: Hutchinson's University Library.

- Plato. 2002. *Five Dialogues: Euthyphro, Apology, Crito, Meno, Phaedo*, 2nd ed. Translated by G.M.A. Grube and John M. Cooper. Indianapolis, IN: Hackett.
- Prichard, H. A. 1912. "Does Moral Philosophy Rest on a Mistake?" *Mind: A Quarterly Review of Psychology and Philosophy* 21 (81): 21–37.
- Quadrio, Philip A. 2009. "Kant and Rousseau on the Critique of Philosophical Theology: The Primacy of Practical Reason." *Sophia* 48 (2): 179–93.
- Rousseau, Jean-Jacques. 1979. *Emile: Or, On Education*. Translated by Allan Bloom. New York: Basic Books.
- Rousseau, Jean-Jacques. 2008. *The Discourses and Other Early Political Writings*. Edited by Victor Gourevitch. Cambridge: Cambridge University Press.
- Schneewind, J. B. 2009. "Kant and Stoic Ethics." In *Essays on the History of Moral Philosophy*, 277–95. Oxford: Oxford University Press.
- Schönecker, Dieter, and Allen W. Wood. 2015. *Immanuel Kant's Groundwork for the Metaphysics of Morals: A Commentary*. Cambridge, MA: Harvard University Press.
- Scott, Dominic. 2006. *Plato's Meno*. Cambridge: Cambridge University Press.
- Sedgwick, Sally. 2008. *Kant's Groundwork of the Metaphysics of Morals: An Introduction*. Cambridge: Cambridge University Press.
- Sensen, Oliver, ed. 2012. *Kant on Moral Autonomy*. Cambridge and New York: Cambridge University Press.
- Shell, Susan Meld. 2009. *Kant and the Limits of Autonomy*. Cambridge, MA: Harvard University Press.
- Shell, Susan, and Richard Velkley. 2017. "Rousseau and Kant: Rousseau's Kantian Legacy." In *Thinking with Rousseau: From Machiavelli to Schmitt*, edited by Helena Rosenblatt and Paul Schweigert, 192–210. Cambridge: Cambridge University Press.
- Sinnott-Armstrong, Walter. 2006. *Moral Skepticisms*. Oxford: Oxford University Press.
- Stern, Robert. 2010. "Moral Scepticism and Agency: Kant and Korsgaard." *Ratio* 23 (4): 451–74.
- Sticker, Martin. 2014. "How Can Common Rational Capacities Confirm the Correctness of the Deduction in Groundwork III — and Why Does It Matter?" *Hegel Bulletin* 35 (2): 228–51.
- Tenenbaum, Sergio, ed. 2010. *Desire, Practical Reason, and the Good*. Oxford: Oxford University Press.
- Thorpe, Lucas. 2006. "The Point of Studying Ethics According to Kant." *Journal of Value Inquiry* 40 (4): 461–74.
- Timmermann, Jens. 2007a. *Kant's Groundwork of the Metaphysics of Morals: A Commentary*. Cambridge and New York: Cambridge University Press.
- . 2007b. "Simplicity and Authority: Reflections on Theory and Practice in Kant's Moral Philosophy." *Journal of Moral Philosophy* 4 (2): 167–82.
- Velkley, Richard L. 1989. *Freedom and the End of Reason: On the Moral Foundation of Kant's Critical Philosophy*. Chicago: University of Chicago Press.
- . 2013. "Transcending Nature, Unifying Reason: On Kant's Debt to Rousseau." In Oliver Sensen, ed., *Kant on Moral Autonomy*, 89–106. Cambridge and New York: Cambridge University Press.
- Velleman, J. David. 1992. "The Guise of the Good." *Noûs* 26 (1): 3–26.
- Ware, Owen. 2014. "Skepticism in Kant's Groundwork." *European Journal of Philosophy* 24 (2): 375–96.
- Wolff, Robert Paul. 1973. *The Autonomy of Reason: A Commentary of Kant's Groundwork of the Metaphysic of Morals*. New York: Harper & Row.
- Wood, Allen W. 1999. *Kant's Ethical Thought*. Cambridge: Cambridge University Press.
- . 2008. *Kantian Ethics*. New York: Cambridge University Press.
- . 2010. "Kant and the Intelligibility of Evil." In *Kant's Anatomy of Evil*, edited by Sharon Anderson-Gold and Pablo Muchnik, 144–72. Cambridge: Cambridge University Press.
- Woolf, Raphael. 2008. "Colloquium 1: Misology and Truth." *Proceedings of the Boston Area Colloquium of Ancient Philosophy* 23 (1): 1–24.
- Zammito, John H. 2002. *Kant, Herder, and the Birth of Anthropology*. Chicago: University of Chicago Press.