Agency and Evil in Fichte’s Ethics

Owen Ware
Simon Fraser University

This is similar to the case of the sailor who preferred to console himself with the hope that he might be able to bear up in hell rather than having had to improve himself in this life. In hell he was merely supposed to suffer, whereas in this life he would have had to do something. — Fichte (SE IV: 202–203).

1. Introduction

Of the many philosophical doctrines Kant set forth in his lifetime, few caused as much scandal and surprise as his claim that all human beings have an innate propensity to ‘radical evil.’ Aside from the fact that it appears contrary to the underlying spirit of his moral philosophy, the most vexing aspect of this doctrine is its glaring lack of argument. After presenting it in his 1793 Religion Within the Boundaries of Mere Reason, Kant says that we can ‘spare ourselves the formal proof that there must be such a corrupt propensity rooted in the human being, in view of the multitude of woeful examples that the experience of human deeds parades before us’ (R 6: 32–33). As commentators have been quick to point out, no amount of empirical evidence can justify a universal claim — that all human beings are prone to evil.¹ So despite what Kant says, it appears the doctrine of radical evil requires a ‘formal proof’ after all.

While this topic has become a point of dispute in recent Kant scholarship, I would like to shift focus in this essay to an earlier figure who was sensitive to the apparent gap in Kant’s argument: J. G. Fichte.²

1. See Michelson (1990) and Bernstein (2002). The objection is not new, though, and Fichte may have encountered it through Johann August Eberhard. In his 1794 essay ‘On Kantian Radical Evil in Human Nature,’ Eberhard wrote: ‘Now how does Herr Kant prove that such a radical evil exists, or that the human being is evil by nature? He does not carry out this proof, as would be expected, from principles of pure reason; rather, it rests merely on experience [bloß auf die Erfahrung]’ (cited in Louden 2011).

2. The Kant scholarship divides between formalist readings (see Allison [1990], [2012]; Morgan [2005]; Palmquist [2008]; Muchnik [2010]) and anthropological readings (see Anderson-Gold [1991]; Wood [1999], [2010]; Frierson [2003]; Sussman [2005]; Louden [2011]; Pasternack [2014]). The difficulty is that Kant is unclear what a ‘proof’ of our propensity to evil would look like, whether it would be transcendental (and so a priori) or anthropological (and so a posteriori). In support of the latter view, Wood (2014) has argued that when Kant speaks of a ‘formal proof’ in the Religion, he is referring to a proof that would result from ‘future’ (weiterhin) research into human nature (R 6:...
In his 1798 *System of Ethics* (hereafter, the *Sittenlehre*), Fichte openly struggles with the question of evil in §16, titled ‘The cause of evil in a finite rational being’; and in the Appendix he seeks to provide the kind of proof Kant mentions, but never delivers, in the *Religion*. Here Fichte observes that mere experience cannot secure the universality of radical evil. ‘There must be some rational ground for this claim,’ he explains, beyond the many examples we have of our corrupt tendencies (SE IV: 199). In the Appendix, Fichte then locates this ground in the fact that all human beings are indolent, subject to what he calls, rather cryptically, a ‘force of inertia’ (SE IV: 199). What is odd, aside from the argument itself, is that Fichte believes Kant draws the same conclusion as he does: that everyone is evil because, quite simply, everyone is lazy. But what basis does Fichte have for making this claim? And why does he associate it with Kant?

In seeking to answer these questions, I shall distance myself from a view accepted by most scholars working in this area today: that Fichte was mistaken to consider his proof of evil Kantian in spirit (Piché [1999]; Kosch [2006], [2011]; Dews [2008]; and Breazeale [2014]). Historically, this view has its source in Schelling (1809) who accused Fichte of losing his nerve in §16 of the *Sittenlehre*, attributing evil to a natural force rather than to an exercise of free will. Against this standard interpretation, I shall argue that the meaning of inertia is figurative: it is a tendency all of us have to prefer passivity over activity, resisting as we do the work of acting for ourselves. In this way the concept of ‘inertia’ is similar to what Kant, in his essay on enlightenment, calls our ‘self-imposed immaturity’ (WA 8: 35). Evil in this sense stems from our unwillingness to be autonomous — in short, it is laziness, but a form of laziness for which we are responsible. Our propensity to resist the moral law is universal, then, because the moral law presents us with the most arduous of tasks: striving, as Fichte puts it, for rational ‘self-sufficiency.’

Contrary to Schelling’s verdict, I do not think Fichte lost his nerve in the *Sittenlehre*, for he did not attribute evil to a force beyond the will. This goes some way toward vindicating Fichte’s self-described Kantianism, although it does not completely defend his argument for evil. As I will show, Fichte’s proof suffers from a missing step. While it explains why all of us prefer passivity over activity, it does not elucidate evil itself, as our act of resisting the moral law. One of my aims in this essay is to reconstruct this missing step in Fichte’s argument using materials he provides elsewhere in the book. Beyond this, my more general aim is to show that despite his self-described Kantianism, Fichte’s ethics is striking for its originality, systematicity, and rigor. What is more, I believe there is a key insight we should take seriously from §16 of the *Sittenlehre*. It is Fichte’s thesis that for embodied agents like us, the authority of reflection — our capacity to step back from ourselves — is far more limited than we might think. Questions of interpretation aside, there is a valuable philosophical lesson we can, and should, learn from this.

2. Preliminary Remarks

What I am calling Fichte’s key insight is easily obscured by his repeated emphasis on the *freedom* we have to raise ourselves up through reflection, or what we may call:

> moral theory and wanted to find this evil that precedes all empirical action in the lethargy of human nature* (*1809: 56)*.
Fichte is nevertheless sensitive to the fact that this process will take time: as embodied agents, we can only raise ourselves up step by step (SE IV: 178). In effect, ‘it will take some time until everything that is originally in us and for us [i.e., our capacity for moral agency] is raised to the level of clear consciousness,’ and Fichte adds that ‘to describe this temporal course of the I’s reflections is to provide the history of an empirical rational being’ (SE IV: 178).

This last remark serves as a clear statement of Fichte’s methodology in the Sittenlehre. Quite unlike Kant, who in the Religion starts from an analysis of a rational agent aware of the moral law, Fichte embraces what we may call a ‘genetic’ account of the human subject. On this account, we begin with what I have called the stage of unreflective agency characteristic of animals and infants, and Fichte’s task is to uncover the higher stages of self-consciousness we can attain over time, ending where Kant begins, i.e., with the perspective of full-blown moral persons. Interestingly, the further thesis underwriting Fichte’s genetic analysis — and the key insight of §16 — is that for embodied agents like us, all acts of reflection are ‘limited’ (begrenzt) (SE IV: 178). Call this:

Limited reflective authority. We cannot suspend the defining features of ourselves — our inner drives or maxims of choice — to the point where they appear normatively arbitrary.7

The point I want to stress here is that while Fichte is committed to moral progress — we are always capable of reflecting on ourselves and ascending to higher levels of agency — he does not think the authority we gain from such reflection is absolute. We cannot step back from our natural drive, for example, to the point where pursuing pleasure (in any of its complex forms) appears optional to us. Reflection opens up space for us to change ourselves, but the change is only ever proximate to the aspect of our agency we reflect upon — whether it be our natural drive to self-preservation or what Fichte calls our pure drive to ‘self-sufficiency’ (Selbständigkeit). For this reason limited reflective authority informs Fichte’s entire approach in §16, namely, to chart the ‘history of an empirical rational being’ (SE IV: 178). The significance of this thesis, however, goes well beyond understanding the complexities of the Sittenlehre. As I shall argue later on, Fichte’s view of our limited reflective powers may resolve a particular skeptical threat facing contemporary Kantian theories of normativity: a threat that concerns why — if at all — we should care about being agents.

3. Four Stages of Agency

For now, let me sketch the four stages of Fichte’s genetic account.

We begin with what I have called the stage of unreflective agency characteristic of animals and infants. At Stage I, a being is wholly motivated by his drive to self-preservation, and he is responsive to this drive in experience, say, by pursuing objects that promote his well-being, and by avoiding objects that hinder it. As spectators we can say this being is ‘free,’ for he is not acting blindly under a mechanical force of instinct. Yet we cannot say he is conscious of such freedom, since he is not aware...
of anything beyond what the natural drive highlights for him. From his own perspective, we can say, he is ‘only an animal’ (SE IV: 178). 

If a being in this state reflects on his natural drive, he will then become conscious of himself ‘over and above’ it. In Fichte’s words, he will ‘tear himself loose from the natural drive by means of this reflection, and position himself as a free intellect independent of [it]’ (SE IV: 179). This will gradually open up new deliberative powers that were previously unavailable to him. At Stage II, he can begin to delay the satisfaction of his natural drive — postponing inclinations that would have otherwise pressed him into action — as well as select among multiple ways of fulfilling those inclinations. Yet in exercising these powers he is still absorbed by the pursuit of pleasure. Since ‘the only drive that occurs within his consciousness at this point is the natural drive, and since the latter aims only at enjoyment’, his highest maxim will be:

Happiness: ‘one must choose that which promises the greatest pleasure, in terms both of intension and of extension’ (SE IV: 180).  

Now if the agent steps back from this maxim — and granting moral progress, there is nothing preventing him from doing so — his attention will no longer be absorbed by the pursuit of pleasure. As a result, he will have just enough room to consider another drive operating within him, what Fichte calls the drive to ‘self-sufficiency’ (SE IV: 185). However, what is characteristic of Stage III, for Fichte, is that an agent

8. For a more elaborate discussion of Fichte’s four stages, see Goh (2012).
9. Fichte adds that ‘one may of course also seek one’s own happiness in the happiness of others, but in this case the ultimate goal of acting still remains the satisfaction of these drives and the pleasure that arises therefrom, and hence one’s own happiness’ (SE IV: 180).
10. Fichte nonetheless has some interesting things to say about how an agent might get stuck at Stage II or Stage III (such as the corrupting influence of bad philosophy) — but these details are not relevant for the discussion at hand.
11. For a broader discussion of self-sufficiency in relation to Fichte’s social and political philosophy, see James (2011).

A final ascent will occur only if the agent steps back and reflects on his drive to self-sufficiency. If he does this, he will see that self-sufficiency is his ‘true essence’ (wahren Wesen) as a rational being — not something that belongs to him contingently or ‘by chance’. With this insight before him, the agent will see that acting according to a concept of self-sufficiency is categorically required of him, that he ought to do so for its own sake. And then, somewhat paradoxically, the same drive that had acquired a blind and despotic character for him — the

Lawless freedom: one must choose that which promises ‘unrestricted and lawless dominion over everything’ (SE IV: 186; italics removed). 

A final ascent will occur only if the agent steps back and reflects on his drive to self-sufficiency. If he does this, he will see that self-sufficiency is his ‘true essence’ (wahren Wesen) as a rational being — not something that belongs to him contingently or ‘by chance’. With this insight before him, the agent will see that acting according to a concept of self-sufficiency is categorically required of him, that he ought to do so for its own sake. And then, somewhat paradoxically, the same drive that had acquired a blind and despotic character for him — the

12. To be clear, the agent we are considering does not deliberately seek unrestricted and lawless dominion. As Fichte puts it, he does not intend ‘to bring everything outside of him under the absolute sway of his will — he does not intend anything at all, but is only driven blindly’ (SE IV: 186). At this stage, the agent’s ‘aim’ — while not consciously worked out — is to remove everything that obstructs his causality (SE IV: 186).
13. This is not something we can do individually, however. Later in the Sittenlehre, Fichte argues that reflection upon our rational self-sufficiency requires the presence of another person, someone who ‘summons’ (Aufforderung) us to free activity (SE IV: 220–222; cf., GN III: 33–36). I have discussed this issue elsewhere (Ware 2010). Relatedly, striving for self-sufficiency is not a private affair, since Fichte believes — in the tradition of Rousseau and Kant — that morality develops ‘by means of purely rational education in the context of social intercourse’ (SE IV: 349).
14. Fichte makes this point earlier: ‘the whole concept of our necessary subjugation to a law arises solely through the absolutely free reflection of the I upon itself in its own true essence, its self-sufficiency’ (SE IV: 56–57). Relatedly, he writes that ‘the law as such becomes a law for the latter when the intellect reflects on it and freely subjects itself thereto, and thus self-actively makes this law into the unbreakable maxim of all its willing’ (SE IV: 56).
pursuit of unrestricted dominion — will undergo a transformation. As Fichte explains this process:

A human being has only to raise to clear consciousness this drive to absolute self-sufficiency — which, when it operates as a blind drive, produces a very immoral character — and then, simply by means of this very act of reflection, this same drive will transform itself within him into an absolutely commanding law. (SE IV: 191)

In the space of reflection, then, what had previously led the agent to seek unobstructed causality will become a law of self-legislation, the law of letting nothing but reason determine his will. So in place of lawless freedom, his highest maxim will be:

*Autonomy:* one must choose that which promotes self-sufficiency, *i.e.*, by letting nothing other than reason influence one’s choices (SE IV: 191).

For Fichte, complying with autonomy requires that the agent act conscientiously. It requires that he act only when he is aware of performing his duty — of legislating himself — instead of letting something or someone else act for him. ‘The whole of moral existence,’ Fichte writes, ‘is nothing other than the continuous self-legislation of a rational being; and where this self-legislation ceases, there immorality begins’ (SE IV: 56; my translation). At Stage IV, we are bound by the autonomy of our reason, and when we act, we are bound to act conscientiously. Behaving any other way, Fichte maintains, would be ‘perverse and malicious’ (*Verkehrtheit und Bosheit*) (SE IV: 156).

4. The Nature of Evil

Before turning to Fichte’s argument for why such perversion is universal, present in all rational human beings, let me clear up two possible sources of confusion. The first has to do with Fichte’s argument that we cannot act against our duty with ‘clear consciousness’ (SE IV: 191). The second concerns his follow-up argument, that we act immorally only by ‘obscuring’ our consciousness of duty (SE IV: 192). Taken at face value these claims suggest Fichte’s view of evil is at odds with Kant’s. But on further inspection, I will argue, they turn out to be unexpectedly close.

4.1. Diabolical Evil

Starting with the first argument, Fichte writes:

It is absolutely impossible and contradictory that anyone with a clear consciousness of his duty [*deutlichen Bewußtsein seiner Pflicht*] at the moment he acts could, in good consciousness, *decide not to do his duty*, that he should rebel against the law, refusing to obey it and making it his maxim not to do his duty, because it is his duty. (SE IV: 191)

At first this passage may sound very un-Kantian, for in the *Religion* Kant defines evil in terms of an agent’s decision to subvert the moral law to self-love. In his view, if the moral law fails to motivate an agent from the standpoint of deliberation, an incentive opposed to it must have influence on the power of choice of the human being in question [...] [Yet] by hypothesis, this can only happen because this human being *incorporates* the incentive (and consequently also...
the deviation from the moral law) into his maxim. (R 6: 24; my emphasis)

On comparing these two passages, it appears Kant is claiming what Fichte says is ‘absolutely impossible’: that a rational human being can decide, if only implicitly, to make exceptions to the law on the occasions it suits him, thereby incorporating ‘deviation from the law’ into the maxim of his will.

Despite these appearances, I am not convinced the two passages conflict. Notice, first, that Fichte only denies the possibility of an agent rebelling against the moral law ‘with a clear consciousness of his duty.’ Such rebellion would require the agent to form a maxim ‘not to do his duty, because it is his duty’ (SE IV: 191). In other words, what Fichte finds contradictory is the concept of a ‘diabolical’ (teuflisch) being, a being who resolves to act contrary to the moral law while he is aware of its normative authority (SE IV: 191). The difficulty is that we are now speaking of a being who, in the moment he recognizes what the moral law demands of him, ‘demands of himself that he not do the very same thing’ (SE IV: 192). ‘At one and the same moment,’ Fichte writes, ‘these contradictory demands would be placed upon him by one and the same power,’ i.e., his power of free will. But that is not possible. Therefore, the concept of a diabolical being ‘annuls itself’ (SE IV: 192).17

Interestingly, Kant constructs a similar argument in the Religion. ‘To think of oneself as a freely acting being,’ he says, ‘yet as exempted from the one law commensurate to such a being (the moral law), would amount to the thought of a cause operating without any law at all,’ and that, he adds, is a ‘contradiction’ (R 6: 35). So Kant concludes: ‘The human being (even the worst) does not repudiate the moral law, whatever his maxims, in rebellious attitude (by revoking obedience to it)’ (R 6: 36). That is to say, when we subvert the moral law to do not have a positive incentive to violate the law; we do not seek evil for the sake of evil. What we do, rather, is invert the proper relationship between the moral law and self-love. We incorporate both into the maxim of our will, but make compliance with the law conditional upon self-love. That is why Kant thinks the best title for human evil is, not devilishness, but ‘perversity’ (Verkehrtheit) (R 6: 36). Fichte, as we have seen, adopts this same terminology in the Sittenlehre.

4.2. Responsibility for Evil

Another possible source of confusion concerns Fichte’s follow-up argument: that we act immorally only by ‘obsuring’ our consciousness of duty (SE IV: 192). After denying that it is possible to will evil for the sake of evil, Fichte writes:

It is, however, quite possible for one to obscure [verdunkle] within oneself the clear consciousness of what duty demands. For such consciousness arises only through an act of absolute spontaneity; and it endures only through the continuation of this same act of freedom. If one ceases to reflect, then this consciousness disappears. (SE IV: 192; modified)

It is tempting to read this passage as evidence that Kant and Fichte part ways on the question of our responsibility for evil.18 Fichte appears to be saying that evil is never something we will. Since our consciousness of duty can never fail to motivate us, as he has argued, the only explanation of how we act immorally is that our consciousness of duty fades. Kant, on the other hand, is explicit in his view that evil

17. Fichte also rejects the intelligibility of willing evil for the sake of evil in The Vocation of Man (cf., BM II: 228).

18. See, for example, Kosch (2011). Kosch argues that there are degrees of moral accountability for Fichte, since she thinks failure to step back and reflect is a form of involuntariness (lack of formal freedom). It follows, in her view, that what results is not evil because we are not responsible for it. But there are two potential problems with this reading. Aside from the fact that Fichte does make that we are responsible for not reflecting (see SE IV: 174 and 181), the reading Kosch defends would leave Fichte exposed to the kind of objection Sidgwick (1907) mounted against Kant — namely, that Kant’s theory commits us to the absurd view that we are responsible only for morally right actions, not morally wrong ones. Thanks to a reviewer of this journal for drawing my attention to these points. See note 32 below.
stems from how we use (or better, misuse) our freedom of will. That is why, for example, Kant argues we cannot trace the source of evil to our natural inclinations (R 6: 34–35). Beyond the fact that such inclinations are morally neutral,

we also cannot presume ourselves responsible for their existence (we cannot because [...] natural inclinations do not have us for their author), though we can well be responsible for the propensity to evil which, since it concerns the morality of the subject and hence is to be found in the latter as a freely acting being, must be capable of being imputed to the subject as itself guilty of it. (R 6: 34–35)

As before, I do not think these quotations express an opposition of views. We have seen that Kant and Fichte both deny it is possible for someone to revoke her consciousness of the moral law’s normative authority (that is why they claim it is impossible for anyone to will evil for the sake of evil). On this point it is safe to say—in contemporary terminology—that Kant and Fichte are both internalists about moral motivation: they think our consciousness of duty is sufficient by itself to motivate us.19 I would argue further that their rejection of diabolical evil is a result of this theoretical commitment. After all, if we accept motivational internalism, the only possible explanation of how we fail to act morally is that a countervailing force worked against us, something external to our consciousness of duty (though not necessarily external to our will).

19. Kant’s commitment to motivational internalism finds textual support in the Religion. After explaining that a ‘human being (even the worst) does not repudiate the moral law, whatever his maxim, in rebellious attitude (by revoking obedience to it),’ he writes: ‘The law rather imposes itself on him irresistibly, because of his moral predisposition; and if no other incentive were at work against it, he would also incorporate it into his supreme maxim as sufficient determination of his power of choice, i.e. he would be morally good’ (R 6: 36; my emphasis). For further discussion, see Tenenbaum (2007).
form of duty ‘as duty.’ Like Adam, we bargain down what the moral law really demands of us.

Looking back, two things should now be clear. First, it should be clear why Fichte argues we cannot act against our consciousness of duty, as that would entail willing evil for the sake of evil. Second, it should be clear why, granting motivational internalism, the only account left for why we fail to act morally is that we are subject to a countervailing force (something external to our consciousness of duty). On these points, I have argued, Kant and Fichte stand in agreement: they both maintain our consciousness of duty is sufficient by itself to motivate us; that the concept of diabolical evil is contradictory; and that human evil, while at bottom mysterious, involves a tendency to rationalize against the moral law’s stringency.20 In their view, instead of revoking obedience to the law, we effectively work to obscure what we know deep down is our duty — that is, we deceive ourselves. Keeping these points in view, we are ready to assess Fichte’s argument that all finite rational beings harbor a propensity to evil, a claim he first introduces in the Appendix to §16 of the Sittenlehre.

5. The Universality of Evil

Unfortunately, this is where we run into major difficulties of interpretation. For in the Appendix Fichte defends the universality of evil on the grounds that everyone is, at bottom, lazy (SE IV: 199). But it is not at all clear what he means by this. According to a long-standing tradition of interpretation, the so-called ‘laziness’ at the root of evil is a product of inertia, where ‘inertia’ (Trägheit) is taken literally as a natural ‘force’ (a vis inertiæ). As we shall see, there are some initial attractions to this reading. But the drawbacks are, I shall argue, too costly.

5.1. Fichte’s Proof: The Standard Interpretation

For recent examples of the standard interpretation, consider the following two remarks:

20. Important differences remain, which I will discuss in section 5 and note 36 below.

As I mentioned earlier, the first advocate of this reading was Schelling, who argued that Fichte locates evil, not in the will, but in the ‘lethargy of human nature’ (1809: 56).22 Given this tradition of interpretation, it

21. For another example, consider the following remark by Kosch: ‘Only if we fail to ponder hard enough to see our duty do we fail to will it. According to Fichte, laziness is “the radical evil in human nature” (though clearly what he has described is not radical evil in the Kantian sense familiar to readers of Religion within the Boundaries of Mere Reason; moral evil in that sense does not exist, for Fichte)’ (2006: 272–273).

22. Piché (1999) argues that Schelling’s view finds support in Fichte’s choice of terms in §16. In this section, he claims, Fichte only uses the expression radical ‘Böse’ twice (both in reference to Kant), preferring radical ‘Übel’ instead (1999: 216). But I do not think this terminological point carries much weight. First, ‘Böse’ appears six times in §16, not two, one of which is in the very title, ‘Über die Ursache des Bösen im endlichen vernünftigen Wesen.’ Secondly, Fichte uses the term ‘Böse’ many times later in the Sittenlehre, each without reference to Kant, and in one he connects ‘Böse’ to rationalization (e.g., making ‘an
is understandable why nearly every scholar working in this area today refuses to take Fichte at his word when he writes, in a spirit of camaraderie, that Kant ‘also makes the same inference’ from evil to inertia (SE IV: 199). If Schelling and his followers are right — and I am going to question this — then Kant and Fichte represent opposite ends of the explanatory spectrum: the former puts evil on the side of Freedom, the latter on the side of Nature.

In all fairness, the standard interpretation appears to have strong textual support. When Fichte first lays out his proof in the Appendix, he begins by discussing how a *vis inertiae* operates in general. A passive object in nature endures, he explains, only because it has a force ‘to remain what it is,’ and this force holds it together, so to speak, giving it unity over time (SE IV: 200). ‘If an opposing force now operates upon it,’ he continues, ‘then it will necessarily resist this with all its force in order to remain what it is; and only now, through its relation to the opposing activity, does what was previously only inertia become an activity’ (SE IV: 200). Fichte then extends this analysis in the following paragraph to encompass human beings. On the assumption that ‘what pertains to nature as a whole must also pertain to the human being insofar as he is nature’ (SE IV: 200), it follows that we are also subject to a *vis inertiae*. Considered as embodied creatures, ‘we ourselves are nothing more than nature’ (SE IV: 200), for nature gives us our first drive, the drive to self-preservation. We too have a tendency to persist in our state and to resist anything that compels us to change our state. Therefore, Fichte concludes, our propensity to resist the moral law is simply ‘inertia or laziness’ (SE IV: 201).

As we can see, the structure of Fichte’s proof is relatively simple:

(1) All things in nature are subject to a *vis inertiae*: the tendency to resist anything opposing their state.

(2) All human beings, as embodied creatures, are parts of nature.

(3) Therefore, given (1)–(2), all human beings are subject to a *vis inertiae*.

Filling in the remaining details, we can say:

(4) The moral law demands that we leave our natural state, *i.e.*, that we strive for rational self-sufficiency.

(5) Therefore, given (1)–(4), all human beings [either (a) can or (b) actually do] resist the moral law.

Without having to settle the disjunctive in Premise 5, we can already see what advocates of the standard interpretation are doing. They are taking the sense of a ‘force of inertia’ in Premise 1 to be the same as the ‘force of inertia’ in Premise 3. That explains why they read the laziness at the root of evil to express a natural power, a power all things display in resisting change. (It also explains why they repudiate Fichte’s self-described Kantianism — for in the Appendix, it seems, Fichte no longer traces evil to human freedom.) Yet we must ask: is it viable to interpret the ‘inertia’ in Premises 1 and 3 as identical? I will argue that it is not. However, aside from giving credit where it is due, I think it will be instructive to see where advocates of the standard interpretation seem to get things right.

5.2. Requirements of a Formal Proof

First, let us review what we know as readers prior to the Appendix. At this point in §16 all we can say for certain is that evil involves obscuring the form of duty as duty, a stratagem of self-deception whose ‘aim’ is to avoid the work of autonomy. Without much controversy, then, we can specify Premise 5 of Fichte’s proof by settling on the first disjunct:

---

exception to the rigor of the law’) (SE IV: 270; see also IV: 288 on the ‘Böse’ of lying). To support his case further, Piché (1999: 217) suggests Fichte’s choice of terms reflects Kant’s distinction between ‘Böse’ and ‘Übel’ from the Critique of Practical Reason (1788). There Kant says ‘Übel’ refers to sensibility (the ‘ill’ of our state), and ‘Böse’ refers to agency (the ‘badness’ of our actions) (*KpV* 5: 60). But clearly, Fichte does not consider the radical ‘Übel’ of human nature an ‘ill,’ which suggests (contra Piché) that he is not straightforwardly adopting Kant’s distinction in §16.

23. Pedro (2006) is a noteworthy exception to this trend, and there is much in her essay with which I agree, especially her critique of Piché (1999).
(5a) All rational human beings can be evil.

For Fichte, however, this statement would be insignificant, even if in some sense true (SE IV: 199). That is, we can accept the possibility of evil, but given the problematic status of this claim, we cannot go further and assert the second, more substantive disjunct:

(5b) All rational human beings are evil.

The reason for this explanatory limit is simple. Clarifying a concept does not amount to establishing its reality. This is a basic Kantian lesson. Yet the lesson presents a further challenge. Establishing a concept’s reality is not always a matter of experience. If a concept bears the mark of universality, for example, no amount of experience will prove that it exists (we do not perceive universality in the way that we see, e.g., color). What this implies, more straightforwardly, is that we cannot defend the truth of (5b)—the existence of evil as a ‘universal human phenomenon’—by observing others. Observing others would get us no further than experience, and experience, as Kant and Fichte both agree, cannot secure universal claims.

Fichte draws these points together in an important passage, just after identifying laziness as ‘a truly positive radical evil’ (SE IV: 199), asking:

And what entitles us to make such a presupposition? Is it merely a matter of experience? This is what Kant seems to assume, despite the fact that he also makes the same inference we are about to make. Mere experience [Erfahrung], however, would not entitle us to make such a universal presupposition. There must therefore be some rational ground [Vernunftgrund] for this claim, though one that does not yield necessity—since that would destroy freedom [indem die Freiheit aufgehoben würde]—but that only explains this universal experience. (SE IV: 199; modified)

Here Fichte is clear about the following: to prove that evil is more than a logical possibility, we need to show that there is an actual basis for it in infinite rational beings. But then we face the difficulty of establishing the reality of evil as a universal phenomenon, since anything universal, as we have just seen, cannot be justified on experiential grounds. For this reason Fichte does not think we should, as Kant recommends, ‘spare ourselves the formal proof’ of evil in light of ‘the multitude of woeful examples that the experience of human deeds parades before us’ (R 6: 32–33). Indeed, Fichte’s point is just the reverse. The only way we can hope to establish (5b) is to offer a formal proof, i.e., a proof that would uncover the rational ground of evil.24

The above passage is also noteworthy for specifying the requirements a ‘formal proof’ must satisfy. In the first place, it must satisfy an Actuality Requirement: it must show that evil has an actual basis in infinite rational beings. Secondly, a formal proof must satisfy a Uniformity Requirement: it must show that all of us (and not just a limited sub-set) are evil. These two conditions are present in the proposition ‘All rational human beings are evil’. There is, however, a third condition Fichte thinks a formal proof must satisfy, though it is less obvious than the first two. It comes up in his last qualifying remark that a proof must uncover the rational ground of evil, ‘though one that does not yield necessity—since that would destroy freedom—but that only explains this universal experience’ (SE IV: 199; modified). Fichte’s claim is that proving the necessity of evil would be too strong; it would mean that

24 Given the contrast he sets up between ‘experience’ (Erfahrung), on the one hand, and a ‘rational ground’ (Vernunftgrund), on the other, Fichte is evidently wanting a ‘formal proof’ of our propensity to evil that does not amount to anthropological study (see note 2 above). If the latter is what Kant ultimately meant by a ‘formal proof’—as Wood (2014) argues—then Fichte is perhaps guilty of misunderstanding Kant’s position in the Religion. I do not think this detracts from Fichte’s analysis in §16 of the Sittenlehre, however, because he is clearly attempting to go beyond Kant in securing the ‘universal presupposition’ of evil. As we shall see, Fichte’s strategy is to say, first, that all of us must work through four stages of agency, and second, that all of us inevitably resist maintaining our clear consciousness of duty at Stage IV. This is the sense in which Fichte’s proof is properly formal. Thanks to a reviewer of this journal for pressing me to clarify this point.
we cannot act otherwise, try as we might. We can then take Fichte to be saying a formal proof must satisfy a third condition, what we may call an Inevitability Requirement: it must show that evil is, not unavoidable, but inevitable for finite rational beings.25

In what way, then, do advocates of the standard interpretation seem to get things right? Returning to Fichte’s proof, recall that they would have us read the ‘force of inertia’ in Premise 3 as natural, identical to the vis inertiae in Premise 1. From this point of view, the standard interpretation is surprisingly effective, at least in two ways. First, it is effective in meeting the Actuality Requirement. We can say evil has an actual basis, since it originates from the very materiality that makes us embodied creatures. For the same reason, the standard interpretation is effective in meeting the Universality Requirement. We can say evil is a universal phenomenon, since it originates from a law governing all objects in the natural world. On this reading, the proposition ‘All rational human beings are evil’ is true because the proposition ‘All things are subject to a vis inertiae’ is true. And that is why the standard interpretation provides a transcendental style of proof after all: in seeking a rational ground for evil, it brings us to an a priori law of nature, nothing that experience on its own could secure for us. That makes it look quite promising.

Nevertheless, when it comes to the Inevitability Requirement, the standard interpretation begins to falter. The problem is as follows. If we say evil originates from the materiality that makes us embodied creatures, we run the risk of making evil necessary for finite rational beings. But then we run the further risk of undermining our responsibility for evil, by making it a natural phenomenon. Fichte warns us about this, as we have seen above. A proof that makes evil necessary would ‘destroy freedom,’ he says, since it would no longer relate to how we exercise our will (SE IV: 199). That is why Fichte stipulates a formal proof must satisfy a third requirement: it must show why evil is to be ‘expected’ (erwarten) but not ‘necessary’ (notwendig), for if evil were necessary we would lack any capacity to act otherwise. While it may sound overly subtle, Fichte’s distinction is important. To satisfy the Inevitability Requirement, he is saying, we must explain the ubiquity of evil without resorting to a fatalistic picture of human agency. Once we acknowledge this, it is clear why the standard interpretation rests on a deceptive selling point: it gains purchase on the Actuality and Universality Requirements only at the cost of making evil unavoidable for us. After all, the basis of its ‘proof’ — Nature’s vis inertiae — is not something for which we could be responsible.

5.3. Fichte’s Proof: An Alternative Interpretation
Fortunately, the task of finding an alternative interpretation is not difficult. As a first step, all we have to do is show that when Fichte speaks of a ‘force of inertia’ in the context of human life (as stated in Premise 3), he is not speaking of a literal force (as stated in Premise 1). In other words, all we have to do is consider the possibility that in the context of human life, the primary meaning of ‘inertia’ is figurative, a concept whose explanatory power works by analogy, rather than by strict identity, to the torpor of natural things. As promised, I shall now pursue this possibility.

In the Appendix, when Fichte remarks that an inanimate object (a rock, for example) has a force to ‘remain what it is,’ it is safe to assume he does not mean a rock has a ‘will’ to remain what it is. Nor is it likely he means a rock has a ‘will’ to resist change when affected by an opposing force (SE IV: 200). This much strikes me as uncontroversial. The sense of ‘force’ (Kraft) in a ‘force of inertia’ (Kraft der Trägheit) is not the same as a human ‘will.’26 I emphasize this because when Fichte extends his analysis to encompass human beings (on the prin-

25. Kant argues along these lines in the Religion. To say ‘A human being is by nature evil’ does not mean we can infer evil from the ‘concept of a human being (i.e.,) from the concept of a human being in general, for then the quality would be necessary’ (R 6: 32; my emphasis). Nevertheless, Kant adds, ‘we may presuppose evil as subjectively necessary in every human being’ (R 6: 32). This captures Fichte’s view, but to avoid confusion, I will speak of ‘inevitability’ instead of ‘subjective necessity.’

26. See Neuhouser (1990) and Zöller (1998) for detailed examinations of Fichte’s theory of will and its development in the 1790s.
Agency and Evil in Fichte’s Ethics

Owen Ware

principle that what pertains to nature as a whole must also pertain to us as parts of nature), we have good reason to suppose a shift of meaning has occurred. Like natural things, we have a tendency to remain in our current state; and like natural things, we have a tendency to resist changing our state. So, like natural things, we also display a kind of ‘inertia’ (SE IV: 200). The crucial difference, however, is that our tendency to remain as we are is something we actively will. I take it this is where the analogy to the torpor of natural things breaks down. We will to remain in our current state; and we will to resist changing our state—and even if these ‘willings’ are not fully conscious, they still express a form of agency entirely lacking in a thing’s inertia. This is enough, in my view, to challenge the guiding assumption behind the standard interpretation.

The case for my alternative interpretation is not strictly conceptual, however. There is good textual support for it in the way Fichte characterizes evil earlier in §16 and elsewhere in the Sittenlehre. First, let us not forget that Fichte claims we are responsible for obscuring our consciousness of duty—to quote him again: ‘it is up to our freedom whether such consciousness continues or becomes obscured’ (SE IV: 192). But we must ask: How could we be free in resisting the moral law, if the resistance was not something we actively willed? Secondly, recall that the most dangerous form of obscurity for Fichte involves rationalizing against the moral law, obscuring the form of duty ‘as duty’ (SE IV: 196). But again, how could we rationalize against the moral law, if we were not freely exercising our will in the process? When we re-examine Fichte’s analysis prior to the Appendix, we find that he defends a view of evil which is very much agency-centered, tied closely to rationalization, deception, and (most centrally) self-deception. Nor is this atypical for the Sittenlehre as a whole. Later in the book when Fichte discusses lying, for example, he writes: ‘The thought of lying requires something positively evil: a deliberate search for some crooked path that will allow one to avoid proceeding along the straight path that offers itself to us’ (SE IV: 288). Such remarks are hard to reconcile on the standard interpretation.

One might object, however, that if we shift interpretive weight to the agency of evil, as I am recommending, we end up contradicting or at least trivializing what Fichte has to say about inertia and laziness in the Appendix. But that verdict would be too hasty. I would like to insist, on the contrary, that putting agency at the foreground of our interpretation provides a key for understanding Fichte’s formal proof. For when we read the Appendix as I am suggesting, the figurative significance of ‘inertia’ and the analogy to nature become clear. To start with, consider what ‘inertia’ could mean from an agency-centered perspective. Obviously it could not be something external to us, a force that literally drags us down from rational exertion. Instead, it must be something internal to us, something we sustain through our will. This is the crux of the metaphor I believe Fichte wants to impress upon his readers: Inertia speaks to a human (all-too-human) desire to be like a thing of Nature. If I am right about this, then the primary meaning of inertia is the psychological inverse of what the standard interpretation takes it to be. It is not a natural force that draws us back into the realm of passivity. Rather, it is an expression of our will, our will to passivity we might say. Instead of trivializing the Appendix, then, an agency-centered perspective in fact explains how inertia manifests in human

27. Goh’s (2012: 11) distinction between ‘freedom of reflection’ and ‘freedom of choice’ may be helpful here. On the one hand, we can say that the stage of agency someone occupies will limit his freedom of choice, in that he will be constrained by whatever ‘way of thinking’ (Denkweise) is characteristic of that stage. On the other hand, we can say that someone (considered as a rational being) still has the freedom to reflect upon himself and ascend to a higher Denkweise. In this way, freedom of reflection pertains to how we move ‘vertically’ between the four stages of agency, whereas freedom of choice pertains to how we move ‘horizontally’ within any given stage. Within this framework, we can understand the agent’s responsibility for evil in terms of his refusal to reflect (an act of freedom, even if not an act of choice, strictly speaking).

28. It is also difficult on the standard interpretation to explain why Fichte would second Kant’s claim that ‘radical evil is inborn in the human being and yet has its ground in freedom’ (SE IV: 182), or why he would later conclude that ‘the evil in human beings has its ground in freedom’ (SE IV: 183).
life. It manifests as our unwillingness to be agents — and that, I submit, is ‘laziness’ in its most generalized form.29

There are, I think, a number of advantages to reading Fichte along these lines. To start with, it helps dispel the air of mystery surrounding his inference from evil to inertia. We desire to be like things in nature, directed by forces beyond our will.30 Morality demands that we be agents, directed by reason alone. More precisely, we desire heteronomy (governance from without), but morality demands autonomy (governance from within). This is why, as strange as it sounds, Fichte thinks ‘laziness’ is the positive root of evil — a point he illustrates with the example of a sailor who took consolation at the thought of going to hell, where he only had to bear up and suffer, rather than remain alive, where he would have to work at improving himself (SE IV: 202–203).

Like the sailor, we prefer anything to the work, and responsibility, of having to act for ourselves. So all of us are prone to obscure our consciousness of morality — to rationalize away its authority — because it demands that we step out of our passivity.31 In this respect, the interpretation I have presented explains, not only why all of us actually resist the moral law (meeting the Actuality and Universality Requirements), but also why this resistance is predictable (meeting the Inevitability Requirement). Faced with what seems to be a Sisyphean task, all of us distort our moral consciousness. Yet since this distortion is something we will, there is no sense in which it is necessary. We are at all times free to act otherwise.32

29. Wood (2015) sees Fichte’s concept of laziness as a precursor to many core themes in 19th and 20th century existentialism, especially to Kierkegaard’s notion of ‘despair.’ While I am unable to explore these connections further, it is worth mentioning that Fichte frames the very decision of adopting a philosophical system in broadly existentialist terms: ‘Whether one embraces or rejects [idealism] is something that depends upon one’s inmost way of thinking and upon one’s faith in oneself. A person who has faith in himself cannot accept any variety of dogmatism orfatalism’ (NM 17; cf., SE IV: 45–26 where Fichte says the ‘entire decision’ of his philosophy lies in affirming the reality of freedom.) See Martin (1997) for further discussion. Also, for an excellent study of Fichte’s influence on Sartre specifically, see Gardner (2005).

30. For Fichte, a version of this desire underlies the position of the ‘dogmatist’ who, by refusing to affirm his own self-sufficiency, affirms the self-sufficiency of things. As Fichte writes: ‘They glimpse their own image only insofar as it is reflected through things, as in a mirror. If they were deprived of these things, then they would lose themselves at the same time. Thus, for the sake of their own selves, they cannot renounce their belief in the self-sufficiency of things; for they themselves continue to exist only in conjunction with these things’ (IWL I: 194). One strength of my account, over the standard interpretation, is that it harmonizes with Fichte’s more general critique of dogmatism in philosophy. The latter, we might say, presents us with a theoretical analogue of our will to passivity.

31. According to the standard interpretation, the claim that inertia inheres in all of nature or matter is a crucial link in the proof: it is because we are natural/material beings that no one is exempt from this force. One might worry, however, that if Fichte only means to highlight the fact that we have a tendency similar to the torpor of natural things — as I am suggesting — then he no longer has any extra-empirical basis for proving that all finite rational beings are evil, and that would fall short of the Universality Requirement. In response to this worry, let me add that no one is exempt from inertia because no one comes into the world as a fully developed rational agent. That means no one is exempt from the work necessary for reaching Stage IV. My interpretation meets the Universality Requirements, then, because our propensity to resist this work (‘inertia’ in the figurative sense) belongs to all finite rational beings. Thanks to Kien-How Goh for pressing me on this point.

32. One may still be puzzled by Fichte’s description of inertia in the Appendix as a force of ‘nature’ which holds us in ‘check,’ or his description of our moral task as one of ‘[stepping] onto the other side of the scale’ to overpower the ‘weight of [our] nature’ (SE IV: 201). At a glance, these remarks appear to support the standard interpretation I have called into question. Nevertheless, it is important to see that while Fichte speaks of ‘weights’ and ‘scales’ (the language of causal mechanism), he writes that ‘[f]rom the point of view just indicated we ourselves are nothing more than nature’ (SE IV: 201), and that ‘[i]f one views this matter in purely natural terms, then it is absolutely impossible that a human being should be able to help himself’ (SE IV: 201). What this reveals is that the naturalistic descriptions we find in the Appendix are the result of a specific explanatory perspective Fichte takes up, the perspective of causal mechanism. Since that is not the only perspective he brings to human action and agency, it would be wrong to maintain that ‘inertia,’ in its core meaning, is nothing more than a force of nature. All of this suggests that proponents of the standard interpretation cannot rest their case solely on passages where Fichte’s language is mechanistic, as with his discussion of ‘weights’ and ‘scales’ at SE IV: 201. They must further show that a mechanistic perspective alone explains how inertia is a positive ground of evil. I do not see how that strategy could work, however.
5.4. Completing Fichte’s Proof
There are two further advantages of my interpretation worth noting.

One is that it lets us add a missing step from Fichte’s proof by explaining the sense in which evil is *perverse*. By tracing evil to generalized laziness, Fichte has identified a ‘rational ground’ for evil that is sufficiently removed from empirical evidence (i.e., from the many examples of human corruption). Yet for all that, he has not elucidated the element of perversity in evil, the sense in which we are conscious of what morality demands of us and yet resort to rationalization all the same. To reconstruct this missing step, let us recall what Fichte says earlier in §16 about the third and fourth stages of agency. A being at Stage III is only obliquely aware of another drive operating within her, the drive to self-sufficiency. For this reason ‘self-sufficiency’ will appear contingent to her, something present in her ‘for no higher reason’ (SE IV: 185). Recall also that Fichte thinks an agent advances to Stage IV only when she realizes, through due reflection, that this drive is not contingent at all, that it is the essence of her rational nature. Then, and only then, will she become aware of autonomy, the demand of letting nothing other than reason determine her will (SE IV: 191). Putting these pieces together, we can say radical evil is perverse because, by making our consciousness of morality obscure, we are really making ourselves obscure—to ourselves. This is not a diabolical attraction to evil for the sake of evil, but a deliberate twisting and warping of what we know on some level is our true nature as moral beings.33

Another advantage of my interpretation is that it sheds light on Fichte’s curious reference to Kant. As we have seen, after delivering his proof in the Appendix, Fichte remarks that Kant makes the ‘same inference’ and argues ‘very correctly’ that all human beings are lazy (SE IV: 202). Most commentators find this remark mysterious, if not misleading, given that they take Kant and Fichte to uphold opposing views of evil. Yet there is another puzzle veiled in this comment. Fichte gives the reader no further hint as to what text of Kant’s he has in mind. On a quick survey, there are only a few passages in his mature corpus where Kant expresses the view Fichte attributes to him, namely, that all human beings are naturally ‘lazy’ (faul).34 Of these passages, the one Fichte likely has in mind comes from Kant’s 1784 essay, ‘An Answer to the Question: What is Enlightenment?’, which famously begins: ‘Enlightenment is the human being’s emergence from his self-imposed immaturity [Unmiindigkeit]’ (WA 8: 35; modified). In the next paragraph Kant says where the root of our immaturity lies: ‘It is because of lazi-ness and cowardice,’ he writes, ‘that so great a part of humankind, after nature has long since emancipated them from other people’s direction (naturaliter maiorennes), nevertheless gladly remains immature for life’ (WA 8: 35; modified).35 Now it is difficult to see what significance this remark could have for Fichte if we read his notion of inertia literally. However, once we read it figuratively, the potency of Fichte’s reference to Kant is easy to grasp. The original laziness all human beings display is like the torpor of natural things—and that is part of its metaphorical

---

33. This is why, for Fichte, an agent who pursues *lawless freedom* at Stage III is not evil: she has no awareness of the moral law. So, there is nothing for her to twist and warp.

34. In their English translation of the *Sittenlehre*, Breazeale and Zöller (2006: 191, note 25) point us to the Fourth Proposition of Kant’s essay on universal history. There he writes that a human being ‘expects resistance everywhere because he knows of himself that he is inclined on his side toward resistance against others. Now it is this resistance that awakens all the powers of the human being, brings him to overcome his propensity to indolence (Faulheit), and, driven by ambition, tyranny, and greed, to obtain for himself a rank among his fellows’ (LS A 8: 21). In my view, we gain a much richer understanding of Fichte’s account of evil by tracing this reference (from SE IV: 202) to Kant’s essay on enlightenment.

35. Similarly, Fichte claims that laziness first produces cowardice. ‘Cowardice is that laziness that prevents us from asserting our freedom and self-sufficiency in our interaction with others’ (SE IV: 202). By way of illustration, he writes: ‘I am terrified by the physical exertion required for resistance, and therefore I subjugate my body; I am terrified by the difficulty of thinking for myself that is inflicted upon me by someone who seems to me to be making bold and complicated claims, and therefore I prefer to believe in his authority in order thereby to rid myself of his demands all the more quickly’ (SE IV: 202). Here, we can once more detect the influence of Kant’s essay on enlightenment. ‘It is so easy to be immature,’ Kant writes. ‘If I have a book to serve as my understanding, a pastor to serve as my conscience, a physician to determine my diet for me, and so on, I need not exert myself at all’ (WA 8: 35).
power—but what the notion really conveys is a propensity we all share to resist acting for ourselves. Transposed in the language of Kant’s essay, inertia is our self-imposed passivity.\(^\text{36}\)

These payoffs aside, a worry may surface that the interpretation I have offered collapses Fichte’s account of evil into Kant’s, and that stands in tension with my larger aim in this essay: to show the originality of Fichte’s thinking in the Sittenlehre. As we know, the view defended by Schelling, and picked up by most commentators since, is that Kant and Fichte represent two sides of a gulf. One locates evil on the side of Freedom, the other on the side of Nature. In rejecting this polarization, have I perhaps reduced any philosophically interesting difference between these two thinkers? I do not think so, mainly because I have not overlooked the fact that Fichte provides the very thing Kant thought was either impossible or unnecessary in the Religion—a ‘formal proof’ of radical evil. That by itself preserves Fichte’s originality, even if the core thesis of his account (the laziness of human nature) builds upon and extends Kant’s notion of ‘immaturity.’\(^\text{37}\) Beyond this, it is worth repeating that my reading of §16 is more charitable than the standard interpretation. When we take an agency-centered view, it is not difficult to see how we are responsible for evil, as well as how evil is a real propensity of will for us. We can then say everyone is disposed to rationalization and self-deception—securing (5b) of Fichte’s proof, that ‘All rational human beings are evil’—without thereby making evil necessary or unavoidable for us. Since our passivity is self-imposed, we are always free to improve ourselves. In this way my account is consistent with moral progress, Fichte’s conviction in the permanent possibility of raising ourselves up through reflection.

### 6. The Authority of Reflection

This last point folds back to the start of my discussion in this essay. As I noted earlier, while Fichte is committed to moral progress, he is also sensitive to the fact that we do not come into the world as fully developed moral persons. Our striving for rational self-sufficiency is very much an achievement—one we attain only after having successfully passed through earlier, non-moral stages of agency. Fichte’s analysis of these stages rests on a key insight, and I believe the value of this insight goes well beyond the bounds of Fichte scholarship. I say this because Fichte’s insight challenges an assumption behind contemporary Kantian theories of normativity: the assumption that as rational beings we enjoy full reflective authority, the capacity to ‘step back’ from our motives and determine the reasons we have for acting on them (see Korsgaard [1996]; Rosati [2003]). In my understanding, Fichte is committed to a different thesis, the view that our reflective authority is always limited, and that however much we step back from ourselves, we cannot suspend our motives to the point where they appear normatively arbitrary. As he says, all acts of reflection are ‘limited’ (begrenzt) (SE IV: 178). This claim deserves our attention, I believe, not only because it offers a promising alternative to contemporary theories of normativity, but also because it gives us resources for solving a skeptical threat those theories end up generating.

The skeptical threat I am referring to is one that comes up in recent attempts to justify the normativity of moral requirements by locating...
them in more basic, ‘constitutive’ features of agency. On one level these accounts share the goal of preempting the skeptic’s traditional question, ‘Why be moral?’, by showing moral requirements to be necessary for agency as such. The following sort of analogies make this idea intuitive. In the same way that the question of why one should build a house that prevents rain from coming in loses force if we explain why doing so is what makes it a house at all, the question of why one should follow moral requirements loses force if we explain why doing so is what makes one an agent at all (see Korsgaard 2009). There are certain standards and aims that are necessary for house-constitution (and preventing rain from coming in is one of them). So, too, there are certain standards and aims necessary for self-constitution (and satisfying moral requirements is one of them). Upon seeing this, the skeptic’s question is supposed to dissolve. Nevertheless, as critics have been ready to reply, all this strategy does is push the question of normativity back. For we can still ask: what makes standards and aims necessary for agency relevant to the point where they deserve the overriding and non-optional status they purport to have? The original worry about what makes moral claims overriding and non-optional has not gone away. It has only re-emerged, now under the guise of the question, ‘Why be an agent?’

How does this question relate to the assumption behind contemporary Kantian theories of normativity? To answer this, consider again the picture of rational agency these theories uphold. The picture consists of a being with the capacity to step back from its motives (broadly construed) and decide from a detached perspective whether it has any reason to endorse them. Reflection, on this picture, gives us authority: whether we accept or reject motives is ‘up to us,’ for motives no longer dominate us (as we may suppose they do for animals and infants).

38. Enoch (2006) has offered the most forceful version of this challenge, aimed primarily at Velleman (2001), Rosati (2003), and Korsgaard (2009). I shall be highlighting one aspect of Enoch’s criticism: his claim that attempts to ground the normativity of moral requirements in aims constitutive of agency fail to explain how those aims can be normatively non-arbitrary. For other criticisms of the constitutivist strategy, see Setiya (2003), Tiffany (2011), and Tubert (2011).

The assumption behind these recent accounts, however, is that we are capable of full reflective authority. On this model, the distance we are able to gain from our motives is so great that they eventually appear normatively arbitrary. And that is where the skeptical threat gets a hook in. For if we are now told that certain capacities are necessary for agency (that without them, we would not count as agents at all), we have every right to ask why we should care about those capacities. Why should the fact that some features of our agency are necessary make a normative difference? Why should acting according to these features be overriding and non-optional — or better, what would warrant us in regarding them that way? The point I want to make is that raising the question ‘Why be an agent?’ becomes possible the moment we buy into the assumption of full reflective authority. The two emerge from the same conceptual space. The assumption behind these recent accounts, however, is that we are capable of full reflective authority. On this model, the distance we are able to gain from our motives is so great that they eventually appear normatively arbitrary. And that is where the skeptical threat gets a hook in. For if we are now told that certain capacities are necessary for agency (that without them, we would not count as agents at all), we have every right to ask why we should care about those capacities. Why should the fact that some features of our agency are necessary make a normative difference? Why should acting according to these features be overriding and non-optional — or better, what would warrant us in regarding them that way? The point I want to make is that raising the question ‘Why be an agent?’ becomes possible the moment we buy into the assumption of full reflective authority. The two emerge from the same conceptual space. The assumption behind these recent accounts, however, is that we are capable of full reflective authority. On this model, the distance we are able to gain from our motives is so great that they eventually appear normatively arbitrary. And that is where the skeptical threat gets a hook in. For if we are now told that certain capacities are necessary for agency (that without them, we would not count as agents at all), we have every right to ask why we should care about those capacities. Why should the fact that some features of our agency are necessary make a normative difference? Why should acting according to these features be overriding and non-optional — or better, what would warrant us in regarding them that way? The point I want to make is that raising the question ‘Why be an agent?’ becomes possible the moment we buy into the assumption of full reflective authority. The two emerge from the same conceptual space. The assumption behind these recent accounts, however, is that we are capable of full reflective authority. On this model, the distance we are able to gain from our motives is so great that they eventually appear normatively arbitrary. And that is where the skeptical threat gets a hook in. For if we are now told that certain capacities are necessary for agency (that without them, we would not count as agents at all), we have every right to ask why we should care about those capacities. Why should the fact that some features of our agency are necessary make a normative difference? Why should acting according to these features be overriding and non-optional — or better, what would warrant us in regarding them that way? The point I want to make is that raising the question ‘Why be an agent?’ becomes possible the moment we buy into the assumption of full reflective authority. The two emerge from the same conceptual space.

While more could be said about these theories of normativity, let us consider how Fichte might respond to the skeptical problem we have before us. A first reply, I imagine, would be something like this. He might say that the very question ‘Why be an agent?’ only gets a hook into our discussion when we assume we enjoy full reflective authority as rational beings. Once we reject this view in favor of Fichte’s thesis — what I have called limited reflective authority — we close off conceptual room for the skeptic’s question to take root. The assumption behind these recent accounts, however, is that we are capable of full reflective authority. On this model, the distance we are able to gain from our motives is so great that they eventually appear normatively arbitrary. And that is where the skeptical threat gets a hook in. For if we are now told that certain capacities are necessary for agency (that without them, we would not count as agents at all), we have every right to ask why we should care about those capacities. Why should the fact that some features of our agency are necessary make a normative difference? Why should acting according to these features be overriding and non-optional — or better, what would warrant us in regarding them that way? The point I want to make is that raising the question ‘Why be an agent?’ becomes possible the moment we buy into the assumption of full reflective authority. The two emerge from the same conceptual space.

39. Enoch’s critique of Rosati (2003) illustrates this problem nicely. In Rosati’s case we are to imagine an agent so reflective that every aspect of her self appears normatively arbitrary. Enoch then asks why this agent should be moved by the discovery that certain aspects of her self are necessary to be an agent at all. From the standpoint of full reflective authority, why should the necessary-for-agency status of such capacities matter? ‘She is, remember, stepping back from her desires, attempting a kind of detached scrutiny, evaluation, and choice [...] But why should it matter, as far as the question of normative arbitrariness is concerned, that some parts of her psychology have this necessary-for-agency status?’ (Enoch [2006: 178]; modified). As I will argue next, my claim is that Fichte’s thesis of limited reflective authority closes off this particular skeptical question. But I am not claiming it closes off every possible objection Enoch could raise.

40. It does not close off the question, ‘Why be autonomous?’ — a point I will return to shortly.
reason for this is clear. If we do not enjoy full reflective authority, then whatever aspects of ourselves we bring into reflection will not appear normatively optional. On this picture, we have no room to ask why we should care about being agents, since there is no point at which the practical relevance of our motives (whatever they may be) fades away. On the contrary, the picture we get from Fichte is one in which our motives retain their import and salience from the standpoint of deliberation. The limited power we have to suspend the force of our inner drives and maxims of choice means they will always appear compelling for us, at least to some degree. To take one example, I can step back and reflect on my drive to self-preservation, but for Fichte my power to reflect only grants me the ability to conceptualize my total happiness, and that in turn opens up new deliberative powers for me (I can, as he says, postpone the satisfaction of my desires, figure out new ways of satisfying them, and so on). But at this stage of agency I only have room to ask how to fulfill my natural drive — not whether I should do so.

If the skeptic’s question has force so long as we assume the standpoint of a fully reflective being — a being for whom all features of her agency appear normatively arbitrary — then the reply I have just sketched would work. All we need to do is reject the particular model of rational agency shared by contemporary Kantian theories of normativity. While I have only touched the contours of an alternative picture, the broader significance of Fichte’s account should now be clear. We can never step so far back from ourselves that our inner drives and maxims of choice lose their normative significance. As we have seen, limited reflective authority is a basic premise of Fichte’s four-stage genetic account. When I step back from my drive to self-preservation (at Stage II), I have just enough space to consider how to fulfill my desires to their fullest extent, and in doing so I bring myself under a new maxim, happiness. When I step back from happiness (at Stage III), I have just enough space to consider another drive operating within me, the drive to self-sufficiency.41 When I bring myself to act under this drive — still half-consciously — I replace happiness with lawless freedom, the pursuit of unlimited dominion over all things. Finally, when I bring my drive to self-sufficiency into the space of reflection (at Stage IV), I discover its law — the law of letting nothing other than reason determine my will. When I bring myself to act under this law — now consciously — I replace lawless freedom with autonomy as the highest maxim of my being. At each of these stages, I am engaged (rather than detached) in my acts of reflection. At no point does the problem of normative arbitrariness arise.

7. Closing Remarks

In this essay I have argued, contrary to the prevailing view in the literature, that Fichte’s proof of evil in §16 of the Sittenlehre does not rest on a natural force of inertia. The concept of inertia is figurative, speaking to a universal tendency we have as finite rational beings to prefer passivity over activity, resisting as we do the work of acting for ourselves. Understood in this way, Fichte’s self-described Kantianism in §16 is not mistaken, despite appearances, for Kant himself openly theorizes about a tendency all of us have to self-imposed ‘immaturity.’ Fichte nonetheless retains his originality in the Sittenlehre, I have argued, since he works out an argument for evil based on this tendency, something Kant never attempted to do.

41. Something like a thesis of full reflective authority is present in Korsgaard’s (1996) account of the ‘normative problem.’ She writes: ‘[O]ur capacity to turn our attention on to our own mental activities is also a capacity to distance ourselves from them, and to call them into question […]’ I desire and I find myself with a powerful impulse to act. But I back up and bring that impulse into view and then I have a certain distance. Now the impulse doesn’t dominate me and now I have a problem. Shall I act? Is this desire really a reason to act? (1996: 93). As we have seen, Fichte believes we can step back from our drive to self-preservation, but not so much that we can ask, ‘Shall I act on it.’ For Fichte, we only have enough space to consider the maximal satisfaction of our desires, both in terms of their intension and extension. Thus, because Fichte thinks acts of reflection provide us only with limited distance, he believes we must work through successive stages of agency.
From everything I have said, however, one may still want to know why an agent at the final stage of Fichte’s account should regard autonomy as normatively superior to lawless freedom. Rather than ask, ‘Why be an agent?,’ one may ask, ‘Why be autonomous?,’ where the latter involves giving priority in matters of deliberation to rational self-sufficiency. In forming a reply, there are two things we should bear in mind. First, an individual at Stage III already aims at something other than the pursuit of pleasure. The fundamental goal around which she organizes her choices concerns a blind and despotic pursuit of freedom, and that carries more practical weight than the fulfillment of her sensible desires. Secondly, we should bear in mind that an individual at this stage regards her drive to self-sufficiency as something contingent or non-essential to her true nature, for at this level she still identifies her true nature with the natural drive. As we have seen, Fichte thinks an individual only reaches the final stage of agency when she realizes, by reflecting on her pure drive, that it expresses the essence of her rational nature. What I would like to say — although Fichte is silent on the issue — is that an individual who reaches Stage IV has no room to question the normative superiority of autonomy over lawless freedom. There is no room for her to ask, ‘Why be autonomous?,’ because autonomy is the essence of her freedom, and that is something she already values more than the pursuit of pleasure.

I suppose Fichte is silent on this issue because he is optimistic about our motivational capacities. He believes that once we recognize an essential feature of our agency — our drive to self-sufficiency, for example — we will be internally motivated to act according to what the drive demands. That is why the problem Fichte struggles with in §16 of the Sittenlehre does not concern normative arbitrariness. To the extent that we recognize the demands of autonomy, he thinks, we will appreciate their practical import over the demands of our natural drive. The problem Fichte struggles with has to do with our will to passivity, that is, our unwillingness to act for ourselves. So the final reply we can imagine Fichte giving to the skeptic who asks, ‘Why be an agent?’ or ‘Why be autonomous?,’ is that these questions may be expressions of resistance: ‘I don’t want to be an agent; I don’t want to be autonomous.’ They may be expressions of laziness. And if that is the case, then what the skeptic needs is not direct argumentation, but something more like inspiration. Such an individual, Fichte says, ‘would have to see himself in his contemptible shape,’ and he ‘would have to see exemplars who elevate him and provide him with an image of how he ought to be, who infuse him with respect, along with a desire to become worthy of respect himself’ (SE IV: 204). A hard task, to be sure; but one we are always free to pursue.

Abbreviations


Kant


Many thanks to Kristin Gjesdal, Anthony Bruno, Yolanda Estes, Kien-How Goh, Dai Heide, Michelle Kosch, Nedim Nomer, Krista Thomason, Evan Tiffany, Leah Ware, Ariel Zylberman, and two anonymous reviewers of this journal for their helpful feedback on earlier versions of this paper.
Agency and Evil in Fichte’s Ethics


Works Cited


Agency and Evil in Fichte’s Ethics


OWEN WARE

