

Kant on Determinism and the Categorical Imperative*

Markus Kohl

I provide a sympathetic reconstruction of Kant's motivation for endorsing incompatibilism about human freedom. On my interpretation, Kant holds that if all the determining grounds of our actions were subject to natural necessity, we would never be free to respect or defy laws of practical reason, and for Kant such freedom is a condition for the possibility that our actions are governed by categorical imperatives. I argue that his view rests on a gripping construal of the rational imperfection that afflicts the human will.

I. INTRODUCTION

Most commentators find little appeal in Kant's seeming incompatibilism about freedom.¹ According to one influential line of interpretation, Kant's incompatibilism is based on a faulty hedonistic conception of empirical motives: Kant worries that if the will were determined by empirical motives, then we would always act for the sake of pleasure, and nonhedonistic moral motivation would be impossible.² For others, Kant's incompatibilism

* For helpful comments on this article, I am grateful to Stephanie Basakis, Zack Bruce, Lara Buchak, Adam Cureton, Andy Engen, Peter Epstein, Jon Garthoff, Hannah Ginsborg, Niko Kolodny, David Palmer, Robert Pippin, Mike Rieppel, Marcus Willaschek, Jay Wallace, and the referees and editors of *Ethics* (especially Talbot Brewer). I also want to thank the Mabelle McLeod Lewis Memorial Foundation for financial support of my dissertation project.

1. Quotations from Kant's works, apart from the *Critique of Pure Reason* which is cited according to the standard A/B pagination, cite the page number of the Academy edition, *Kants Gesammelte Schriften*, ed. Königlich Preussische Akademie der Wissenschaften, vols. 1–29 (Berlin: de Gruyter, 1902–). Translations are my own. I will use the following abbreviations: Man = *Metaphysical Foundations of Natural Science*; GMS = *Groundwork of the Metaphysics of Morals*; MS = *Metaphysics of Morals*; KpV = *Critique of Practical Reason*; KU = *Critique of Judgment*; Rel = *Religion within the Bounds of Pure Reason*.

2. See Terence Irwin, "Morality and Personality: Kant and Green," in *Self and Nature in Kant's Philosophy*, ed. Allen Wood (Ithaca, NY: Cornell University Press, 1984), 31–56, 39–40;

Ethics 125 (January 2015): 331–356

© 2015 by The University of Chicago. All rights reserved. 0014-1704/2015/12502-0001\$10.00

bilism is not based on any actual argument: it reflects a mere ‘dogmatic attachment’.³ The conviction that Kant lacks convincing grounds for deeming causal determinism incompatible with freedom has led still others to suggest that Kant is proposing an ingenuous form of compatibilism.⁴

In this essay I aim to show that Kant has a nondogmatic motivation for rejecting compatibilism about free will or free agency,⁵ a motivation that has not been fully appreciated by commentators.⁶ My primary aim is interpretive: after arguing against a widespread compatibilist interpretation, I try to identify the grounds of Kant’s incompatibilism by focusing on his conception of the possibility of objective *oughts* (categorical imperatives). But I also suggest that Kant’s view rests on interesting philosophical ideas that bear on contemporary debates about the extent to which norm-governed agency requires the possibility of wrong action.

I will use the term ‘determinism’ to denote a view that affirms two claims: (1) Every event is caused by what Kant calls ‘empirical’ or ‘natural’ causes (these causes operate according to deterministic natural laws and produce their effects with nomological necessity); (2) There are

Allen Wood, “Kant’s Compatibilism,” in Wood, *Self and Nature in Kant’s Philosophy*, 73–101, 82–83.

3. See Karl Ameriks, *Kant’s Theory of Mind* (Oxford: Oxford University Press, 2000), 227.

4. See H. E. Matthews, “Strawson on Transcendental Idealism,” *Philosophical Quarterly* 19 (1969): 204–20; Jonathan Bennett, *Kant’s Dialectic* (Cambridge: Cambridge University Press, 1974), 211–23, and “Commentary: Kant’s Theory of Freedom,” in Wood, *Self and Nature in Kant’s Philosophy*, 102–12; Thomas Hill, *Dignity and Practical Reason in Kant’s Moral Theory* (Ithaca, NY: Cornell University Press, 1992), 135–40; Christine Korsgaard, *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996). I should clarify that when I speak of a ‘compatibilist interpretation’, I mean a purely compatibilist reading. Some (e.g., Wood, “Kant’s Compatibilism”) suggest that Kant’s doctrine is a hybrid of compatibilism and incompatibilism. This complication is irrelevant here because the hybrid reading also raises and faces the question of why Kant rejects a pure compatibilism.

5. In Kant there is no tight distinction between freedom of will or freedom of action because the exercise of the will (in choices of maxims) is itself an “inner” action (MS, 218).

6. Henry Allison traces Kant’s incompatibilism to the idea that agents exercise “a certain control over” their desires by incorporating them into their maxims, i.e., by forming a rule which states what desires are worth acting on. See Henry Allison, *Kant’s Theory of Freedom* (Cambridge: Cambridge University Press, 1990), 39. This ‘incorporation thesis’ is clearly an important part of Kant’s doctrine. However, it cannot by itself explain Kant’s incompatibilist commitments because it is not clear why this thesis could not be understood in purely compatibilist terms: Why should Kant be unhappy with the idea that acts of incorporation are free insofar as they are causally determined by the agent’s reflection on reasons? It may seem especially puzzling why Kant would not allow that we are free when our reflection causally determines us to choose the right maxim (e.g., not to act on a desire when doing so would be impermissible). After all, as Allison notes, Kant’s conception of deterministic natural causes allows “for a rich and potentially attractive form of compatibilism” (ibid., 34) that includes among the determining causes not just desires but also beliefs, intentions, and representations of reasons (KpV, 96). This is an important point, for it suggests that Ameriks is too hasty when he faults Kant “for not giving a fair chance to compatibilism.” See Karl Ameriks, *Interpreting Kant’s Critiques* (Oxford: Oxford University Press, 2003), 178.

no further (what Kant calls) ‘nonnatural’ causes of events (i.e., causes which operate outside the deterministic order of nature). It is arguable that Kant rejects (2) and posits a nonempirical, nondeterministic (‘noumenal’) form of causality.⁷ But for the purposes of my article, this point is mostly irrelevant. My goal in what follows is only to explain why Kant endorses incompatibilism about freedom.⁸ I do not try to show how he defends the libertarian view that we are free in a sense that requires the falsity of determinism.

II. A COMPATIBILIST READING

I will first sketch a contemporary form of compatibilism that claims to be inspired by Kant, before examining the interpretive claim that Kant himself accepts this compatibilism. On the compatibilist view I have in mind, our freedom requires that we act on the basis of deliberative choice, and determinism appears to threaten freedom via the predictability of our actions: adopting the deliberative stance requires that one takes oneself to have deliberative options, an attitude that would be precluded by a predictive belief that one will perform a particular act.⁹ Proponents of this view argue that this threat is necessarily empty. Suppose someone (myself or a third party) tries to predict how I will act. My learning of the prediction may well be causally relevant to how I decide to act (e.g., I might have counter-predictive motives). If so, the prediction is among the data that the predictor must use for calculating the prediction; thus, the attempt to calculate the prediction presupposes, *per impossible*, that the prediction is known before it is calculated.¹⁰ If, on the other hand, the prediction has no bearing on my action, the prediction must not interfere with my deliberative choice qua cause of my action, and this requires that the prediction is kept secret from me or that I ignore (i.e., disbelieve) it.¹¹ So

7. For a seminal defense of this point, see Wood, “Kant’s Compatibilism.” The current essay abstracts, as far as possible, from Kant’s idealist distinction between appearances and things in themselves. For my interpretation of how Kant’s views on freedom relate to his idealism, see Markus Kohl, “Kant on Idealism, Freedom, and Standpoints,” *Archiv für Geschichte der Philosophie* (forthcoming).

8. In this article I do not discuss Kant’s incompatibilism about moral responsibility, which raises separate (albeit closely related) issues.

9. See Hilary Bok, *Freedom and Responsibility* (Princeton, NJ: Princeton University Press, 1998), 79; Christine Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996), 95–96; Gilbert Ryle, *The Concept of Mind* (London: Hutchinson’s University Library, 1949), 196.

10. See Bok, *Freedom and Responsibility*, 81–85; Korsgaard, *The Sources of Normativity*, 95–98; Ryle, *The Concept of Mind*, 196–97.

11. Arguably, I cannot deliberate about whether to do x if I believe that I will do x. See Jay Wallace, *Responsibility and the Moral Sentiments* (Cambridge, MA: Harvard University Press, 1994), 3–4.

predictions of my action are either impossible or irrelevant to my practical deliberation.

This argument seeks to show that determinism cannot undermine the alternative possibilities needed for freedom. The operative notion of ‘possibility’ here is epistemic: for a deliberating agent, different actions are “possible for all [she] could possibly know.”¹² If I am confronted in my deliberation with a theoretical prediction of how I will act, it is up to me to decide the practical relevance of this information: thus (*pace* fatalism), if determinism is true, my deliberative decision making remains an indispensable condition of agency that prediction cannot bypass or preempt. Hence, whatever fact there may be about how one is determined to act is inaccessible and (thus) irrelevant to someone trying to decide what is worth doing.

The *interpretive* claim that Kant endorses something like this compatibilist view is rather popular.¹³ It is primarily based on passages in *Groundwork III*, where Kant stresses that the idea of free will arises only from within a practical standpoint and that we are “free in a practical respect” (GMS, 448). According to the compatibilist interpretation, this means that whether we really are causally determined is irrelevant to our practical stance: as Christine Korsgaard puts it, “[Kant’s] point is not that you must believe that you are free, but that you must choose as if you were free. . . . Kant’s answer to the question whether it matters if we are in fact . . . free [from natural necessitation] is that it does not matter.”¹⁴ But Kant’s emphasis on the practical character of our commitment to free will might also be understood as follows: while deliberators must believe that their will is free from natural necessitation, they should acknowledge that we cannot theoretically comprehend such freedom. The idea of freedom cannot be employed from within the standpoint of theoretical reasoning because it cannot answer to theoretical purposes such as explaining the metaphysics of free will (GMS, 459–60; KpV, 133).¹⁵

12. Bok, *Freedom and Responsibility*, 120; cf. G. E. Moore, *Ethics* (Oxford: Oxford University Press, 1912), 94.

13. I list a number of commentators who suggest this reading in n. 4 above.

14. Korsgaard, *Creating the Kingdom of Ends*, 162, 176.

15. I wholly agree with Hill (*Dignity and Practical Reason in Kant’s Moral Theory*, 136) and Korsgaard (*Creating the Kingdom of Ends*, 203) that Kant sees no point in pursuing metaphysical queries about how free will works. They identify an important sense in which Kant’s account differs from contemporary libertarian views that seek to theoretically analyze and explain the metaphysics of free will. By contrast, Eric Watkins reads Kant as offering a view akin to contemporary theories of substance causation that purports to explain why free acts happen at the time they do. See Eric Watkins, *Kant and the Metaphysics of Causality* (Cambridge: Cambridge University Press, 2004), 301–61. To my mind, this reading flies in the face of Kant’s admonition that “we can explain nothing but what we can reduce to [natural] laws” (GMS, 459).

To assess whether the compatibilist reading is right in denying that Kant deems the belief that we really are free from causal determination important to practical deliberation, we must scrutinize its two main contentions: (1) Kant's diagnosis of the problem about the compatibility of free will and determinism concerns the threat that theoretical prediction might deprive us of options ("[predictive] knowledge could somehow take away our freedom").¹⁶ (2) Kant's solution draws on the insight that theoretical predictions cannot preempt our practical deliberation.

Concerning (1), Kant indeed denies that we can predict human action. But this point comes up exclusively in the context of his philosophy of science, when he argues that empirical psychology cannot cognize the psychological laws that would enable predictions about our motive causes (Man, 471–72). In his discussion of freedom, Kant leaves this view unmentioned and even imagines that our actions are fully predictable (A550/B578; KpV, 99). This strongly suggests that his worry about free will and determinism is (*pace* [1]) not driven by concerns about predictability. For if that were Kant's worry, it would be inexplicable why his discussion of free will ignores his conviction that our epistemic grasp of psychological motive causes is (in principle) too impoverished to support predictions about how we will act.

Concerning (2), I conjecture that if Kant did seek to save free will from the threat of determinism by emphasizing that the truth of determinism cannot take away our deliberative options and cannot preempt our deliberative efforts, then he would indeed concede that it does not matter whether our deliberation and choice are in fact causally determined. He would then endorse the compatibilist idea that an act can be called free if it results from the agent's deliberative choice, regardless of whether her deliberation and choice are causally determined. But instead, Kant dismisses precisely this compatibilist view (according to which an act is free if its "determining natural ground" is a "representation thought through reason"; KpV, 96) as a "wretched subterfuge" that only gives us the "freedom of a turnspit" (KpV, 95–96). For Kant, it is the (putative) fact of complete causal determination that raises a problem for the freedom we attribute to ourselves from the practical standpoint: "if every event [were] determined by another in time, in accordance with necessary laws," this would "involve the elimination of all practical freedom" (A534/B562). Hence, the compatibilist reading cannot be sustained.¹⁷

16. Korsgaard, *The Sources of Normativity*, 95.

17. One might wonder why Kant could not at least accept compatibilism in the following sense: (1) as deliberating agents we must regard ourselves as free from causal determination, and therefore (2) we are bound by whatever normative law would bind a truly metaphysically free being, regardless of whether we indeed are free from causal determination. This proposal faces several problems. First, it is unclear how the proposal

III. THE GROUNDS OF KANT'S INCOMPATIBILISM

Why, then, does Kant think that determinism is incompatible with a type of freedom that we must presuppose as deliberating agents? I suggest that we should focus here on a passage where Kant declares that “ought has no meaning whatsoever when one has only the course of nature in view” (A547/B575). For Kant “the course of nature” as such is deterministic (A533/B561). So, when he imagines that one has “only the course of nature in view,” what he envisages is a world in which all activity results from deterministic causes, that is, in which the exercise of every capacity “itself stand[s] under another cause determining it in time” (A533/B561) so that “all causality . . . [is] mere . . . nature” (A534/B562). He claims that oughts do not apply in such a world: “it is impossible that in [nature] anything ought to be different from the way it actually is in those [deterministic] time relations,” and thus “we cannot ask: what ought to happen in nature?” (A547/B575). In an exclusively deterministic world, oughts would be meaningless and our behavior would lack the normative dimension provided by meaningful oughts.¹⁸

Why does Kant think this? Here we can begin by noting that Kant assumes that deterministic causes render their effects (nomologically) necessary (A201–2/B246–47): in an exclusively deterministic world, every cause necessitates its effect(s) in accordance with natural laws (A534/B562). Now let us consider how, in such a world, deliberative recognition of norms relates to acts of compliance or noncompliance. Importantly, Kant does not assume that determinism would render mental and volitional states (e.g., beliefs and intentions) causally inefficacious: rather, he holds that if such states cause actions, then they (like all other empirical

could lend support to compatibilism: Why would one need to regard oneself as free from causal determination unless one perceived, in an incompatibilist vein, causal determination as a threat to one's freedom? (I discuss this point at length in Kohl, “Kant on Idealism, Freedom, and Standpoints.”) Second, I could entail only that we must regard ourselves as being bound by laws that would bind truly free beings but not that we are indeed bound by such laws. On Kant's account, we can infer that we are free from causal determination given our consciousness that we are bound by moral laws (KpV, 4, 29–30). Correspondingly, on his view we cannot just presuppose morality while bracketing the metaphysical issue of determinism: “the mechanism of nature” threatens both morality and freedom (Bxxix).

18. Derk Pereboom also discusses A547/B575, but he does not explain why Kant makes these claims about ‘ought’. See Derk Pereboom, “Kant on Transcendental Freedom,” *Philosophy and Phenomenological Research* 73 (2005): 537–67. Further, Pereboom assumes that on Kant's view ‘oughts’ would be false rather than (as Kant says) meaningless in a deterministic world. I conjecture that for Kant, causally determined agents are not the types of beings about which one can make true or false ought claims. It is only with regard to beings who are appropriate subjects for ought-claims that we can draw a distinction between true and false ‘oughts’. (‘One ought to lie whenever this benefits’ is a meaningful, false claim about free agents.)

causes) causally determine their effects (KpV, 96–97). Thus, if our deliberative choices lead us to act, they causally determine our acts of compliance or noncompliance with norms.

Now assume that it is impossible to do something other than what one is causally determined to do. (I will discuss this assumption soon.) If so, our practical deliberation never leaves us the option ‘to go either way’ regarding normative laws (i.e., to accept or defy such laws). By *normative laws*, I mean principles that give necessary reasons for acting and that apply to every rational agent. For Kant the only normative practical laws are moral norms (GMS, 416; KpV, 20–21), which we apprehend as *categorical imperatives*. On my reading, Kant holds that such categorical imperatives, representing practical laws, govern our actions only if we have the real option to accept or defy them. I believe that Kant is led here by two commitments.

First, Kant holds that normative (as opposed to natural) laws apply only to agents who have the capacity to respond to the reasons provided by those laws (GMS, 412). If someone is truly incapable of correctly responding to normative reasons, that is, if she is causally necessitated not to comply with a valid norm (because her reflection on norms is completely inefficacious or because it causally determines her to go against reason), then the norm fails to govern her behavior. To use a famous metaphor, her behavior belongs exclusively to the space of natural causes rather than to the space of norms and reasons.

Second, Kant sees an irreducible difference between the way in which perfect (or holy) and imperfect wills are influenced by their recognition of normative laws. In the case of imperfect wills, the ability to respond to the reasons provided by normative laws has an ontological ‘shadow’ or privation (MS, 227), namely, the propensity to violate these laws. Due to this propensity an imperfect “will is not by its nature necessarily obedient” to laws of reason (GMS, 413). While a perfect will’s recognition of the rational force of normative laws leaves that will with no option but to do what the law says must be done, an imperfect will’s recognition of the right reasons cannot inevitably lead to right action because such a will faces alternatives or obstacles to right action (MS, 394). Given the presence of these alternatives, the rational influence of normative laws cannot make it impossible for finite agents to violate those laws: the idea that imperfectly rational agents stand under normative laws presupposes the option of noncompliance. When an imperfectly rational deliberator recognizes the force of normative reasons, any rational influence would leave her with the option to go against those reasons: “inevitable” determination or necessitation by reason is the privilege of the Holy Will (GMS, 412–13). If a finite deliberator’s attendance to normative laws does inevitably determine her to act and hence deprives her of alternative options, it cannot be reason that determines her: she is

necessitated by nonrational factors that belong to the space of natural causes.

One might object that this point can easily be accommodated by assuming that an imperfect will's appreciation of norms can only bring that will close(r) to right action: the will's compliance with norms occurs only when further causal factors (other than recognition of the normative force of reasons) become operative. But Kant would protest that in this situation (as proposed by the objector) the agent's rational powers are construed as impotent rather than (merely) imperfect: since the agent's appreciation of the right reasons cannot by itself move her to act, her action does not fully accord with those reasons and (thus) falls short of genuine compliance with the law that represents those reasons. In Kant's terms, such an agent lacks the ability to *act from duty*. Acting from duty requires that "the thought of duty be of itself a sufficient incentive" (MS, 393), that is, that an agent's rational knowledge that an action is unconditionally good (KpV, 74) or practically necessary (GMS, 412) can move her to act without depending on further (nonrational) causal ingredients.¹⁹ For Kant, the presumption that we can be sufficiently moved by our knowledge of the good, or that we possess the rational powers required for truly lawful agency, is central to our deliberative self-awareness.²⁰

Thus, for Kant the deliberative plight we face as finitely rational creatures is that the recognition of normative laws can by itself suffice for lawful agency but also leaves us the option to defy laws of reason. The difference between the way in which perfect and finite wills are influenced by their recognition of normative laws comes to light precisely in the fact that only the latter apprehend such laws in the form of *oughts*. In the case of a divine being whose "volition is already by itself necessarily in unison with the law," "the ought is . . . out of place" (GMS, 414). The ought "represents the practical rule in relation to a will which does not imme-

19. For finite agents, the thought of unconditional goodness or practical necessity is the same as the thought of duty or obligation (GMS, 412–13, 439). See the nn. 20 and 21.

20. Hence his frequent appeal to the idea that pure practical reason can of itself determine the will (KpV, 15, 19, 28, 42). Does this refute my earlier point that determination by reason is the privilege of the Holy Will? No, because these are two different notions of determination: only determination of the Holy Will by reason is "inevitable." In the case of imperfect wills, "actions which are recognized as objectively necessary are subjectively contingent," and "the determination of [our] will according to objective laws is a constraint" (GMS, 412–13). The notion of determination of the will via practical constraint by reason is equivalent to the concept of acting from duty, which is inapplicable to the Holy Will (GMS, 434, 439). Notice also that Kant locates the 'determination' of a will by reason "in the intelligible [noumenal] order of things" (KpV, 48), which differs from the natural course of events that is governed by deterministic laws (GMS, 452; A540–41/B568–69). Hence: "The ought expresses a kind of necessity and connection with reasons that does not take place in the whole of nature" (A547/B576).

diately perform an action simply because it is good" (GMS, 414).²¹ Hence, the function of the categorical *ought* is to represent the right reasons to agents who may choose the right or the wrong thing: such freedom is "the ground of the possibility of categorical imperatives" (MS, 222).²² Kant believes that in a deterministic world human agents are always deprived either of the option to act rightly or of the option to act wrongly. Hence, he infers that in such a world categorical imperatives are impossible.

More formally put, Kant's incompatibilism rests on his acceptance of two claims: (i) If one is under an obligation to act in a certain way, then one can act otherwise than one in fact does. (ii) If one is causally determined to act in a certain way, one cannot do otherwise. For Kant, (i) and (ii) show that causal determinism is incompatible with the kind of freedom that allows imperfectly rational agents to be governed by practical laws.²³

Kant accepts (ii) only as a consequence of the supposition that 'all causality is mere nature': if all causes of our actions are deterministic, then we cannot act differently than we in fact do. Kant here mostly relies on the intuition that we cannot act contrary to the influence of deterministic causes and nomological necessity unless we have certain capacities whose exercise is removed from such influence. However, Kant's discussion of compatibilism suggests a response to one influential compatibilist attempt to avoid (ii), namely, to the conditional analysis of 'can': R can do x if (and only if) R would do x if she chose to do x. For

21. Robert Stern provides an illuminating account of the general contrast between holy and finite wills. See Robert Stern, "Kant, Moral Obligation, and the Holy Will," in *Kant on Practical Justification*, ed. Mark Timmons and Sorin Baiasu (Oxford: Oxford University Press, 2013), 125–52. However, he falsely equates the concepts of 'obligation', 'duty', and 'ought' with the idea of practical *necessity* (ibid., 127–29). A Holy Will would cognize morally right actions "as objectively [practically] necessary" (GMS, 412), but it would represent this necessity through a 'be' rather than an 'ought' (KU, 403).

22. One may object here that Kant rejects the idea that we can define free will "through the capacity to choose to act for or against the law" (MS, 226–27). See Douglas Lavin, "Practical Reason and the Possibility of Error," *Ethics* 114 (2004): 424–57, 448. But this remark must be understood within its proper dialectical context. At MS, 226–27, Kant is reacting against the proposal that we can define and explain free will through the experience of wrong action. Kant rejects this approach because he considers free will a nonempirical capacity (i.e., an unobservable, theoretically inexplicable power) that cannot be understood through a definition that draws on experience. My analysis respects this constraint. My appeal to a nondeterministic influence of reason on the will intends to capture our practical self-conception but not to provide a theoretical understanding of free will (see n. 15 above). Moreover, my account is not based on the experience (in Kant's technical, spatiotemporal sense) of wrong action but on an a priori consideration of the different ways in which laws of reason influence perfect and imperfect wills. Kant himself argues, through a priori reflection on the concept of evil, that the propensity to violate practical laws "is . . . deeply rooted" in man's will (Rel, 35).

23. For Kant determinism is compatible with a 'lesser' kind of freedom—namely, the (as Kant calls it) "psychological" freedom to satisfy one's desires. I explain this point in Sec. VI.

Kant, determinism implies that the choices that determine our physical acts are themselves part of the deterministic chain of time and nature; as such they are themselves causally determined by preceding states. Now suppose R chooses not to do x. If R's choice not to do x is causally determined, then it seems that R cannot choose to do x.²⁴ But if R cannot choose to do x, then the fact that R would do x if she chose to do x fails to show that she really can do x.²⁵ Thus, Kant's complaint that the compatibilist analysis is a "petty word-jugglery" (KpV, 96) rests on more than mere dogmatism.

Contemporary compatibilists have come up with more refined attempts to deny (ii). But a speculation of how Kant would respond to all these proposals is beyond the scope and ambition of this article. The more interesting question, in my view—and some compatibilists would agree here²⁶—concerns Kant's precise grounds for accepting (i), that is, for claiming that human agency under normative laws requires that agents can accept or defy the right reasons.

Claim (i) has two subcomponents. First, ought-governed agency requires the ability to act for the right reasons; second, such agency requires the option to go against reason. The first point is perhaps the more intuitive one. Kant assumes that if an agent is truly incapable of responding to the reasons provided by a normative law, there is no relevant sense in which the law *governs* her action: for an action to be governed by a normative law of reason, the agent must be 'within the space of reasons'; that is, she must be capable of understanding and correctly responding to those reasons. Many philosophers, including compatibilists, would be sympathetic to this claim or to the cognate principle 'Ought implies can' (OIC).²⁷

24. As Kant puts it, "internal" and "psychological" causes, conceived as parts of the course of nature, stand "under the necessitation of conditions of past time, which therefore, when the subject has to act, are no longer in his power" (KpV, 96).

25. To make this vivid, we can suppose that a phobia determines R not to choose x. See Keith Lehrer, "Cans without Ifs," *Analysis* 29 (1968): 29–32. In response, the compatibilist might appeal to a further conditional: roughly, R can choose to do x if R would choose to do x if she chose to choose to do x. See Moore, *Ethics*, 94. But this invites an infinite regress, i.e., "a proliferation of conditionals and a proliferation of objections." See Susan Wolf, "Asymmetrical Freedom," *Journal of Philosophy* 77 (1980): 151–66, 154.

26. For Wallace (*Responsibility and the Moral Sentiments*, 223), compatibilists give away the game when they accept the significance of alternative possibilities because they "are then compelled to rest the case for compatibilism on distinctions between different senses of opportunity . . . that seem too fine-grained and technical to do the normative work required of them."

27. Compatibilists who accept OIC include Dana Nelkin and Jay Wallace. See Dana Nelkin, *Making Sense of Freedom and Responsibility* (Oxford: Oxford University Press, 2011); Wallace, *Responsibility and the Moral Sentiments*. OIC is not uncontroversial: for recent criticism, see Peter Graham, "Ought and Ability," *Philosophical Review* 120 (2011): 337–82. Note,

But it may seem puzzling why Kant would endorse the second sub-component of (i), that is, the view that ought-governed agency requires that the addressee can go against reason. I explore this issue in the following two sections.

IV. TWO CONCEPTIONS OF RATIONAL IMPERFECTION

As Douglas Lavin has shown, the notion that governance by norms requires the possibility of violation is accepted by many contemporary philosophers.²⁸ However, the vague idea that finite agents ‘can’ go against reason may be interpreted in different ways.²⁹ For Kant, our rational imperfection is so radical that we must possess the real option to go against reason whenever we act under practical laws: we could never be inevitably determined by reason to act rightly. I will refer to his view as ‘Strong Imperfection’.³⁰

though, that Kant accepts OIC only within important limits. Contemporary debates focus on the idea that moral ‘ought’ implies the ability and opportunity to perform specific physical acts. See Graham, “Ought and Ability,” 341–42; Ishtiyaque Haji, *Deontic Morality and Control* (Cambridge: Cambridge University Press, 2002), 21–22. But for Kant moral ‘ought’ implies no more than the volitional power to choose the right motive in response to the correct (justifying) reasons (GMS, 394; KpV, 36–37). This is due to Kant’s conception of morally right action. For Kant such action must exhibit a good will (roughly: it must display respect for persons). This requires that one earnestly strives to produce certain physical effects but not that one succeeds in doing so.

28. See Lavin, “Practical Reason and the Possibility of Error,” 424–25, for relevant citations.

29. See *ibid.*, 426–29.

30. Kant’s thesis of ‘Strong Imperfection’ may seem similar to contemporary views such as the one proposed by Ishtiyaque Haji. See Haji, *Deontic Morality and Control*, and *Reason’s Debt to Freedom* (Oxford: Oxford University Press, 2009). Haji argues as follows (*Deontic Morality and Control*, 29–30): if one accepts (1) ‘Morally ought implies can’, one must also accept the corollary (2): ‘If one morally ought not to do A, one can refrain from doing A’. Now Haji treats ‘morally ought not’ as conceptually equivalent to ‘wrong’. He thus moves from 2 to (3): ‘If it is wrong for one to do A, one can refrain from doing A’. From 3, Haji moves (34–35) to the further claim (4): ‘If it is right for one to do A, one can refrain from doing A’. Since Haji treats ‘right’ as conceptually equivalent to ‘morally ought to’, he moves from 4 to (5): ‘If one morally ought to do A, one can refrain from doing A’. Is 5 not the same as Kant’s thesis of Strong Imperfection? To see why not, consider Haji’s crucial inference from 3 to 4. Haji’s inference is based on his principle (P): ‘If concepts (‘ought’, ‘right’, ‘wrong’) are in the same deontic family, then, barring a good reason to think otherwise, they have the same freedom condition’ (see *Reason’s Debt to Freedom*, 25, 50). Kant, however, rejects (P). His reason for rejecting (P) is the conceivability of a Holy Will, i.e., of a morally perfect agent with regard to whom “the ought is out of place” because he lacks the option to act wrongly, but who (*pace* 4) acts rightly despite having no option to do otherwise. Accordingly, Kant does not treat (i) ‘ought’ and ‘ought not’ as conceptually equivalent to (ii) ‘right’ and ‘wrong’: for Kant (i) and (ii) differ with regard to what options they entail because they (potentially) apply to different types of free agents. Hence, Kant

But one may also understand the ‘possibility’ to go against reason in a weaker sense that undercuts the incompatibilist implications of Kant’s position. According to what I call ‘Weak Imperfection’, we can honor the fact that oughts can meaningfully be addressed only to imperfectly rational agents without treating the possibility of violation as a condition on all individual ought-governed action: we can allow that finite agents can be inevitably determined by reason to act rightly. Even when this is allowed, we can specify two senses in which it is possible for finite agents to go against reason. First, the notion of ‘possibility’ may be interpreted diachronically: even if finite agents are on particular occasions causally determined to comply with normative laws, one may account for their imperfection by considering their agency across time, that is, by picking out instances of noncompliance which attest to the fact that they sometimes have the option to act wrongly. Second, the notion of possibility may be interpreted in terms of conceivability. Even if our recognition of a normative law determines us to comply with the law, the necessity here is hypothetical: compliance with a norm that is caused by deterministic processes depends on contingent factors. We can always coherently conceive that these factors might be (or have been) otherwise and (thus) that we might act (or have acted) wrongly. By contrast, God’s compliance with practical laws is unconditionally necessary, and therefore his noncompliance is, unlike ours, inconceivable. This, one might say, suffices to distinguish imperfectly rational agency from divine agency under normative principles.³¹

In this section, I will expound two ‘Kantian’ worries about Weak Imperfection: worries that, while not articulated by Kant himself, draw on Kant’s conception of the human condition. In the following section I identify why, precisely, Kant accepts Strong Imperfection.

and Haji understand and defend the idea that ‘ought’ implies the option to act wrongly in entirely different ways. In Haji’s argument, the notion of rational imperfection plays no role at all.

31. While ‘Weak Imperfection’ is not explicitly advocated in the literature, it seems to me implied by a widely held view. That view grants the “trivial” (Lavin, “Practical Reason and the Possibility of Error,” 441) conceptual fact that ‘ought’ entails the possibility of violation but denies that agents governed by ‘oughts’ must have the ‘liberty of indifference’ to accept or defy the right reasons. Thus, on this view, the notion of ‘possibility to act wrongly’ contained in the concept of ‘ought’ must be weaker than the notion that agents under imperatives have the real option to act wrongly in every case of ‘ought’-governed action. ‘Weak Imperfection’ tries to capture what this weaker notion amounts to. The view I have in mind is elegantly articulated by Lavin, *ibid.*, 441–49. Lavin presents his view as a defense of Wolf’s “Reason View,” according to which the ability to act in accordance with reason is sufficient for freedom. See Susan Wolf, *Freedom within Reason* (Oxford: Oxford University Press, 1990), 61–62. A weak conceptual sense of the ‘possibility’ to go against reason is also suggested by John McDowell’s appeal to a (mere) “potential gap” between ideal and actual motivation. See John McDowell, *Mind, Value and Reality* (Cambridge, MA: Harvard University Press, 1998), 105.

The first Kantian worry about Weak Imperfection is that it downplays the impact that our rational imperfection has in cases where our appreciation of norms leads us to choose correctly: Weak Imperfection allows that if an imperfectly rational deliberator acts rightly because of her recognition of the relevant norm, this recognition removes her option to choose the wrong thing. This conflicts with the idea that even in those cases where our deliberation leads us to choose correctly, the conditions that render our rationality imperfect (such as proneness to temptation) inveigh against the rational influence exerted by our recognition of normative laws. Weak Imperfection allows that in such cases, the conditions that seek to divert us from choosing correctly have no claim on us: they are rendered powerless by our recognition of normative laws. One might say that the influence of these conditions is not negated if the right choice occurs as the result of a process of deliberative competition between various motives. But the notion that the factors that account for our rational imperfection genuinely compete against our power to choose for the right reasons requires that these factors have a chance of success (i.e., that they may sway us to choose the wrong thing). If it is impossible for us to choose the wrong thing, there is no such chance of success. The concession that our rational imperfection has no chance of successfully interfering with our responsiveness to reasons conflicts with the idea that for imperfectly rational agents, responsiveness to reason involves struggle against the temptation to deviate from laws of reason (KpV, 32, 74–75).

One might wonder in what sense a competing desire has more influence in a case where one chooses freely not to act on the desire than in a case where one is causally determined not to act on the desire. In response, it can be said that a desire whose presence yields a genuine option to deviate from reason affects (and afflicts) deliberative choice in a deeper way than a desire that one cannot act on. To see this, consider how the fact that one is causally determined not to act on a tempting desire affects the validity of the deliberator's self-awareness. A deterministic causal process may involve, as parts of the causal chain, events that cause the psychological experience of temptation or struggle. But there is, arguably, something illusory about the experience of struggle or temptation in a case where one is causally necessitated not to act on the tempting consideration: it seems essential to genuine temptation and struggle that what one is tempted by, or what one struggles against, can sway one to act. Similarly, it seems constitutive of the experience of genuine temptation that one takes oneself to be aware of a real possibility that the tempting motive can sway one to act. Consider a paradigm case of temptation: Tom feels drawn to cheat on his beloved wife, despite recognizing that doing so would be wrong. Here Tom's experience of temptation revolves around a lively representation that he may choose to

give in to his carnal desire (i.e., that this option is genuinely open to him). If Tom lacked that representation, he (arguably) would not conceive of himself as being engaged in a struggle against a genuine threat of deviating from reason. Moreover, if Tom manages to resist the desire to cheat due to his awareness of the practical law that declares cheating to be wrong, his act of subordinating his desire to reason is, as Kant emphasizes, accompanied by a sense of constraint (KpV, 32, 80). Acting under the awareness of constraint implies that one represents oneself as renouncing an attractive option that one in fact possesses: hence, if it is not possible for Tom to act on his desire to cheat, there is something deeply misleading about the awareness of constraint that accompanies his choice to subordinate the bidding of his desire to the law of reason.

Now, why could the defender of Weak Imperfection not be happy to account simply for the illusory experience of temptation or constraint (in cases where one is causally determined to choose correctly)? Weak Imperfection was offered in an attempt to show that doing justice to the idea of rational imperfection does not require Kant's 'radical' interpretation. If Weak Imperfection entails (via its allowance that imperfect wills can be inevitably determined to act rightly) that characteristic aspects of our practical experience of agency are illusions, this suggests that Kant's conception of rational imperfection is more adequate to our deliberative self-conception.

There is a further Kantian worry about Weak Imperfection. As we saw, Weak Imperfection locates our imperfect rationality in two features: in acts of noncompliance across time and in the fact that our compliance is only ever hypothetically necessary. Now, notice that the first of these features cannot be definitive of our imperfection (it can only illustrate it): the extent to which we are noncompliant across time is itself contingent, whereas we take our rational imperfection to be a necessary aspect of the human condition. Hence, it is only the second feature that allows Weak Imperfection to specify an idea of imperfection that applies to any finite agent as such. Now imagine Sara, who, due to her place in the causal nexus, is always causally determined (via her deliberative choice) to comply with normative laws. To be sure, Sara is still distinguishable from God: we can imagine scenarios in which she would be susceptible to real temptation because the factors (e.g., her psychology) that in the actual world necessitate her to compliance are different. But Sara's case makes vivid that Weak Imperfection allows that depending on how the world happens to go, the presence of real options to deviate from reason may completely vanish from human choice. This implication of Weak Imperfection conflicts with Kant's idea that the propensity to evil is a deep structural feature of the human will whose grip on a free human agent cannot be reduced to merely hypothetical relevance by the contingent course of events (Rel, 32–33). Relatedly, the sense in which

Sara can still be considered imperfectly rational seems anodyne, precisely because her imperfection has been turned into an object of idle counterfactual speculation that does not afflict her actual powers of choice or her practical self-awareness.³² Weak Imperfection cannot account for the intuition that there is something missing here, something which deliberators who are aware of real options to deviate from laws of reason deem integral to their self-conception as imperfectly rational agents. Since Weak Imperfection defines rational imperfection in terms of the mere conceivability that agents may act wrongly, it must hold that Sara's agency exhibits the essence of rational imperfection. If we find that Sara's agency does not accord with our intuitive sense of our rational imperfection, this tells against Weak Imperfection.

V. THE MORAL PSYCHOLOGY OF RATIONAL IMPERFECTION

Even if the considerations expounded in the preceding section suggest that Weak Imperfection is too weak, they do not positively identify those structural features of finite agency that motivate Kant's more radical interpretation of our rational imperfection: features which account for our pervasive real option to deviate from laws of reason. In this section I explain what kind of moral psychology underlies Kant's endorsement of Strong Imperfection, and I argue that Kant's account can be defended against some important objections.

Here we can begin by considering a mundane example of law-governed action. Suppose I see someone misplace one thousand dollars. If Strong Imperfection implied that every finite agent must overcome a temptation to keep the money, it would rest on a false psychological generalization. However, Kant's basis for endorsing Strong Imperfection is that for every finite agent it is always possible to act on the basis of some one of the empirically given desires that affect finite beings: this may be (i) a greedy desire to keep the money or (ii) a desire to return the money that is fueled by a fear of being watched or (iii) a desire to return the money that is fueled by the want that others be happy (etc.). If one acts on the basis of (ii) or (iii), one's act of returning the money contains the letter but not the spirit of the law (KpV, 71–72; Rel, 30).

But this provokes a further question: Why does Kant posit this opposition between acting on the basis of empirical desires and acting

32. One cannot respond here, in the vein of David Lewis, that invoking 'nearly possible worlds' in which Sara acts wrongly shows that she has the real option to act wrongly: the whole point of Weak Imperfection is that we can use the counterfactual possibility of wrong action to replace the 'radical' idea that human agents possess real options to act wrongly. For Lewis's puzzling 'local miracle compatibilism', see David Lewis, "Are We Free to Break the Laws?" in *Free Will*, ed. Gary Watson (Oxford: Oxford University Press, 2003), 122–29.

on the basis of practical laws? Some argue that Kant's idea that acting on a benevolent desire such as (iii) displays a rationally deficient motivation derives from his view that acting on the basis of moral reasons is the only way to counteract our striving for pleasure.³³ If this is true, then the unsympathetic 'hedonistic' interpretation of Kant's incompatibilism (which I mentioned in Sec. I) is right after all: Kant's endorsement of Strong Imperfection, and thereby his incompatibilism, rests on the hedonistic notion that all empirical desires (including [iii]) aim at pleasure.

Whether Kant is indeed committed to such hedonism is a matter of debate.³⁴ But I want to suggest that we can sidestep this debate here, for Kant's endorsement of Strong Imperfection is not (primarily) driven by concerns about hedonistic motivation. Let us grant that my desire (iii) does not have as its object the pleasure I get from making others happy but (solely) their happiness. This desire, qua empirical state, depends on my contingent psychological makeup, that is, on my "love for people and affectionate benevolence" (KpV, 82), emotional states that cause me to identify with the feelings of others and that make it hard for me to bear their (impending) distress. Now, normative laws provide reasons for acting that apply to every rational agent. So the validity and authority of the law that prescribes returning the money is independent of whether the addressee happens to have benevolent desires. Thus, if my choice to return the money is based on my affectionate want that others be happy, it is not responsive to the objective reasons for returning the money that are represented by practical laws: my motive here is "only subjectively valid and merely empirical, and . . . lack[s] the necessity which is represented in every law, namely, an objective necessity arising from a priori reasons" (KpV, 26). For Kant, a motive that aims at satisfying a contingent empirical desire cannot coincide with (i.e., capture the spirit of) a motive that is based on the recognition of necessary reasons, because those reasons apply to every rational agent and so the motives they supply must differ from the motives supplied by desires that rational agents may or may not have.

To illustrate this point, we can notice two things. First, if the motivating ground of my benevolent acts is my affectionate want that others be happy, then all it would take for me not to act benevolently is a lapse of affection, due to (say) a sudden depression that dampens my desire to make others happy and thus interferes with the motivational grounds of my benevolent acts. So my willingness to make others happy strictly rests on egocentric ('subjective'), nonrational states and circumstances

33. For this view, see (among others) Philippa Foot, "Morality as a System of Hypothetical Imperatives," *Philosophical Review* 81 (1972): 305–16; Irwin, "Morality and Personality," 39–40; Wood, "Kant's Compatibilism," 82–83.

34. See Allison, *Kant's Theory of Freedom*, 102–3.

such as the development of my hormone level. Second, even assuming my benevolent psychology remains as it is, my willingness to choose on the basis of my affection for others may well prompt strictly illegal physical acts (i.e., acts that do not even contain the letter of the moral law; *Rel*, 30).³⁵ Suppose a lazy student explains the terrible consequences that failing the class would have for her and asks, with tears in her eyes, for a rewrite of the final. If my affectionate desire that others do not suffer constitutes the motivational ground of my choices, I will (impermissibly) grant the rewrite. Cases like these illustrate Kant's insistence that responsiveness to intersubjectively shared moral reasons cannot be simulated by motives which depend on subjective psychological conditions such as contingent desires, even if those desires happen to be benevolent ones (*KpV*, 19).³⁶ The other side of this coin is that a free agent whose empirical psychology happens to be such that the suffering of others leaves her emotionally cold can comply with the demands of morality (*GMS*, 398).

Kant's basic point here can be summed up as follows. (1) Every finite agent is always (via her sensible nature) affected by empirical desires that impel her to choose on their behalf (i.e., to make their satisfaction a condition of choice).³⁷ (2) Practical laws give necessary reasons for acting that are not based on empirical desires. (3) Thus, every finite agent is always affected by desires that impel her to choose contrary to the spirit of practical laws.³⁸

This argument provides a principled motivation for Strong Imperfection. Strong Imperfection says that we always have the real option to go

35. These two points illustrate why empirical inclinations cannot be (as Kant puts it in his lectures on ethics) the "measuring-rod" or the "mainspring" of moral motivation.

36. However, Kant does not hold that if (in the previous example) I return money to its owner for the right reasons, my moral worth is increased if I must struggle against a desire to keep the money or decreased if I also want to see the owner happy, as long as my act is not based on this desire. See Barbara Herman, *The Practice of Moral Judgment* (Cambridge, MA: Harvard University Press, 1996), 12, 21. (But see also nn. 42 and 46 below.)

37. "However, we find our nature as sensible beings constituted in such a way that the matter of the faculty of desire . . . first presents itself to us; and our pathologically affected self, although it is in its maxims quite unfit for universal legislation, yet, just as if it constituted our entire self, strives to put its pretensions forward first, and to have them acknowledged as the first and original" (*KpV*, 74).

38. "If a rational creature could ever reach this point, that he thoroughly likes to do all moral laws, this would mean that there does not exist in him even the possibility of a desire that would tempt him to deviate from them; for to overcome such a desire always costs the subject some sacrifice and therefore requires self-compulsion, i.e. inward constraint to something that one does not wholly like to do. But no creature can ever reach this stage of moral disposition. For since he is a creature and therefore always dependent with respect to what he demands for complete satisfaction, he can never be quite free from desires and inclinations: and as these rest on physical causes, they can never of themselves coincide with the moral law, the sources of which are quite different" (*KpV*, 83–84).

against reason. The above argument traces our omnipresent propensity to act wrongly to the fact that we are persistently affected by desires that incline us to choose on their behalf. This identifies the sources of Strong Imperfection without invoking a hedonistic conception of nonmoral motives: Kant's point is that all our empirically given desires, regardless of their object, incline us to choose on their behalf and hence impel us not to choose according to necessary reasons whose validity is independent of the possession of those very desires. That our empirical desires impel us toward their satisfaction and thereby provide us with motives for choice seems plausible: to have a desire just means that one feels attracted or drawn (to some degree) toward realizing its object.

I now want to clarify the implications of Strong Imperfection by discussing a few worries that one might raise against it. First, one might hold that the above argument only shows that we are always prone to fail to be moved by the spirit of practical laws. Why should such failure indicate that one acts contrary to that spirit? In response, it can be argued that a failure to be moved by the spirit of practical laws signals a negative attitude (disrespect or lack of care) for that spirit. After all, such failure is not a passive occurrence that merely happens to a free agent: for Kant, such failure results when we actively take a stand on our evaluative priorities, that is, when we grant precedence to our empirical desires and thus subordinate the weight of necessary reasons to the "pretensions" (KpV, 74) of those desires (Rel, 36–37). This shows something about our moral character, namely, a (maybe implicit) disrespectful or careless attitude toward the values represented by practical laws. For instance, one cannot fail to be moved by the humanity of persons without disrespecting (however surreptitiously) the value of humanity.³⁹

Another, rather common complaint is that Strong Imperfection implies that we can always "make a choice for any principle" and that a human agent "has the ability to act against everything he believes in and everything he cares about."⁴⁰ This objection rests on a misunderstanding. While it is true that on the account I am ascribing to Kant human agents always have the option to deviate from laws of reason, our alternatives to acting for the right reasons are delimited tightly by the bounds of our empirical psychology (i.e., by our contingent desire base which partly determines what we care about).⁴¹ Thus, Kant can concede (e.g.)

39. Kant's claim that a failure to be moved by (the spirit of) practical laws implies a moral fault is connected to his rejection of the Stoic concept of *adiaphora*: for Kant there are no morally indifferent free actions, i.e., actions that stand in no relation to the moral law (Rel, 22–24).

40. See Lavin, "Practical Reason and the Possibility of Error," 447, for the first quote and Wolf, "Asymmetrical Freedom," 153, for the second quote.

41. On Kant's view, agents also care about the necessary ends of morality (GMS, 457–58). Wolf, "Asymmetrical Freedom," 152–54, claims that it is incoherent to say that agents who act rightly are not psychologically determined and that they could do otherwise. She

that for someone blessed with natural affection for children, it is (psychologically) impossible to ignore the plea of a helpless injured child. But for Kant there remains the question of whether her choice is based on her contingent benevolent feelings toward children or rather on her rational recognition of the value that the child has as a person, that is, of the norm that prescribes that one must help the child whether or not one happens to have affectionate feelings toward children.

Resistance toward Strong Imperfection also comes from the Aristotelian tradition in which virtue enables a perfect harmony of reason and desire. The conflict between Kantian and Aristotelian moral psychology raises complex issues that I cannot discuss here; I want to suggest only one fundamental reason why Kant rejects the Aristotelian view. To account for the harmony of reason and desire, one must suppose that the object of desire may completely coincide with the object of moral laws of reason. For Kant, this is impossible because contingently given desires whose presence and strength rests on subjective conditions cannot impel us to act for objective reasons whose validity is independent of our possession of those very desires (KpV, 34).⁴² Now, Aristotle's view is not without response to this worry. For Aristotle, there are objective, noncontingent facts about what our desires really aim at:⁴³ these desires

assumes that an agent can act on the basis of the right reasons only if he is determined by "his interests" which "are determined by his heredity or environment." This assumption seems based on the subjectivist idea (which Kant rejects) that what counts as 'the right reason' for R depends on R's empirically given interests.

42. My interpretation here revalidates the traditional (e.g., Hegelian) idea that in Kant's ethics sensibility (empirical desire or inclination) is essentially opposed to morality. This idea has been denied in recent attempts to bring Kant's ethics closer to feeling and (thereby) to reduce the gap between Kantian and Aristotelian ethics. For instance, Allen Wood suggests that Kant can allow for "cases where action for [moral] reason[s] is easy and natural, harmonizing with empirical desires." See Allen Wood, "Kant on Practical Reason," in Timmons and Baiasu, *Kant on Practical Justification*, 57–86, 57. Similarly, Philip Stratton-Lake suggests that Kant's emphasis on the contingent connection between inclination and duty expresses "the fact that the moral law places a rational limit on the practical possibilities open to us in certain circumstances" but not "a conception of inclination as essentially opposed to morality." See Philip Stratton-Lake, *Kant, Duty and Moral Worth* (London: Routledge, 2000), 37–38. However, for Kant this contingency is significant precisely because it highlights the insurmountable gap between moral reasons that are "objective [and] valid for the will of every rational being" and the subjective, private motives furnished by empirical inclinations that are "valid only for the will of the [particular] subject" (KpV, 19). Motives that have merely subjective validity are essentially (i.e., by their very content) opposed to objectively valid moral reasons.

43. We can attribute to Aristotle something close to Kant's concept of 'empirical desire'. Aristotle posits desiderative states that belong to the "irrational element in the soul," i.e., to the "appetitive and... the desiring element" (EN, 1102a32–1103a3). He further accepts that these desires are contingent on (what we would call) empirical factors, such as upbringing (EN, 1103b6–25, 1114a4–21): these factors determine how one subjectively conceives of 'happiness' (EN, 1095a12–25, 1097b1–23). I will use the abbreviations EN for Aristotle's *Nicomachean Ethics* and EE for Aristotle's *Eudemian Ethics*. Quotes are drawn from *The Complete*

find their proper satisfaction, and we achieve the universal human goal of ‘happiness’, only when we follow the voice of practical reason and exercise the virtues (in a complete life; EN 1098a15–20). But Aristotle’s view here depends on the idea that man has a natural telos that determines what the human good (our ‘happiness’) and the satisfaction of our natural desires consists in (EN 1097b22–23). Kant rejects this idea: for Kant, what makes people happy depends on their subjective preferences (KpV, 20–21).⁴⁴

To confirm that Aristotle’s teleological conception of the human good is integral to his defense of the perfect harmony of objective rational prescriptions and empirical desires, we can notice that some prominent Neo-Aristotelians who leave behind Aristotle’s teleology guarantee such harmony by making the prescriptions of practical reason depend on the agent’s subjective desire base.⁴⁵ Hence, Kant would challenge the Aristotelian to specify how the ends of our empirically given desires can coincide with the ends of morality without (on the one hand) sacrificing the objectivity and necessity of moral norms and without (on the other hand) relying on questionable (‘premodern’) teleological assumptions about happiness or the human good. Of course, nothing I have said here implies that this challenge cannot be met.

Kant departs from Aristotle by adopting a ‘modern’ conception of virtue as continence (i.e., as “moral disposition in struggle”; KpV, 84) against our contingent desires.⁴⁶ This invites the worry that Kant’s view

Works of Aristotle, vol. 2, *The Revised Oxford Translation*, ed. Jonathan Barnes (Princeton, NJ: Princeton University Press, 1984).

44. The disagreement here is not merely verbal. For Aristotle as for Kant, ‘happiness’ is the ideal of a life that is pleasant on the whole (EE, 1249a18; EN, 1153b14–18; KpV, 22). Moreover, both think that our conception of happiness is grounded in what we find pleasant (EN, 1104b4–17; EE, 1227a31–1227b12; KpV, 23). Now, Aristotle holds that a vicious agent is confused both about what gives her true pleasure and about what happiness consists in; consequently, the vicious agent’s life is bound to be wrecked by pain and grief (EN, 1166b13–29), whereas an agent who (continuously) exercises the virtues is immune against misery (EN, 1100b33–34) and derives ‘true pleasure’ from her activities. By contrast, for Kant the claim that virtue yields the pleasure required for happiness is false as a judgment of experience (KpV, 114), and the pleasures of bad people, just like the misery of (apparently) good people, are data of experience.

45. I have in mind here chiefly Foot and Elizabeth Anscombe, who both reject the idea that there can be necessary laws of reason. See Foot, “Morality as a System of Hypothetical Imperatives”; Elizabeth Anscombe, “Modern Moral Philosophy,” *Philosophy* 33 (1958): 1–19. By contrast, John McDowell treats moral requirements as categorical imperatives that, in the case of virtuous agents, ‘silence’ competing motives. See John McDowell, “Are Moral Requirements Hypothetical Imperatives?” *Proceedings of the Aristotelian Society* 52 (1978): 13–29. But this is puzzling: the Aristotelian virtuous agent altogether lacks competing motives and hence (as Anscombe notes) she does not represent ethical norms as imperatives that constrain her activity.

46. Many deny that for Kant the struggle against inclination is a mark of virtue or moral worth. See Allison, *Kant’s Theory of Freedom*, 114; Marcia Baron, *Kantian Ethics Almost*

implies, implausibly, that our deliberative activity is typically fraught with the experience of struggle against temptation. However, struggle against temptation requires a reflective awareness of the potential usurpation of our grounds of choice by unlawful motives, and Kant does not claim that this awareness accompanies our everyday activity. On Kant's view, human agents do not typically reflect on their reasons for acting because they typically act on the basis of maxims (KpV, 74), volitional patterns that include general policies to act in certain ways and that foreclose reflection on specific practical questions. Moreover, Kant holds that the extent to which agents become aware of (1) a temptation to choose on grounds that are contrary to the spirit of the moral law is oftentimes contingent on the extent to which they become aware of (2) a temptation to act contrary even to the letter of the moral law,⁴⁷ and he thinks that many agents are, because of luck and risk aversion, which they mistake for virtue, rarely confronted with a temptation of the latter sort (MS, 392–93). Hence, the idea that it is always possible for finite agents to choose contrary to the spirit of practical laws does not entail that this possibility is always an object of conscious awareness, in the form of temptation or struggle. Rather, for Kant, the most typical manifestation of our rational imperfection is self-deception about our motives that derives from a want of reflection on the grounds of our choices and that makes us confuse unlawful habitual patterns of choice (that do not prompt illegal acts merely due to fortuitous chance) with virtue (Rel, 38).

One might propose that the appeal to self-deception suggests a moderation of Strong Imperfection: instead of positing the metaphysical possibility that we may always choose contrary to the spirit of practical laws, why not say that it is always (merely) epistemically possible for us to choose incorrectly, that is, that we can never be certain that our particular acts or general maxims are truly lawful? Kant accepts this epistemic idea (MS, 392), but only as the epistemic upshot of the metaphysics of

without Apology (Ithaca, NY: Cornell University Press, 1999), 199; Herman, *Practice of Moral Judgment*, 32. This view is a reaction against the claim that Kant adopts a 'battle citation' model of moral worth. For this claim, see Richard Henson, "What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action," *Philosophical Review* 88 (1979): 39–54, 48. I agree that on Kant's view agents need not typically struggle against temptations to violate the letter of the moral law. But, as we saw, for Kant we have a pervasive tendency to adopt grounds of choice that violate the spirit of the moral law. And, for Kant it is the defining mark of human virtue that we incessantly engage our sensible inclinations in battle to overcome the "obstacles" that confront our attempts to respect the spirit of practical laws (MS, 394, 409).

47. To illustrate: my maxim to tell the truth may have been based, all along, on fear of bad consequences (in which case it violated the spirit of the law), but my reflective awareness of this may be sparked only when I find myself 'blessed' with an occasion where no such consequences loom and where I feel tempted to lie (i.e., to violate the letter of the law).

imperfectly rational agency: for Kant, the ubiquitous doubt about whether we succeed in respecting the spirit of practical laws derives from the ubiquitous metaphysical possibility that we may fail to respect the spirit of practical laws. One may object that even if it is not always metaphysically possible for us to act on the basis of inadequate motives, we can always doubt the adequacy of our motives because we do not know whether this metaphysical possibility obtains in a given case. But this strikes me as an inconsequential position. Doubts about whether our grounds of choice are pure or corrupted seem idle unless they can be motivated. Kant motivates these doubts by appeal to the possible influence of empirical desires on our grounds of choice (GMS, 419). If the presence of such desires provides a metaphysically real basis for the possibility to choose incorrectly in one case, and if we are always (to some degree) drawn to act on the basis of those desires, then this metaphysical possibility of choosing incorrectly generalizes to all cases (of acts that may be governed by normative laws). For Kant, the omnipresence of empirical desires is guaranteed by the fact that we are dependent beings; as such, we are always affected by some contingent ‘want’ that, again simply as such, draws us toward its satisfaction and whose influence cannot coincide with the spirit of a necessary law of reason.

It should be noted that for Kant our rational imperfection is not only a matter of self-deception but may also (albeit not standardly) manifest itself when a person reflectively confronts temptations to violate good principles of choice that she sincerely holds and, assuming that the temptation is not resisted, when she consciously violates such principles (GMS, 455; Rel, 20). Suppose I feel tempted to lie so that I may avoid a very embarrassing confession. Kant’s view allows us to interpret this as a case in which my desire to escape shame inveighs against my principled conviction that it is wrong to subordinate the dignity of others to my personal gratification. That I sincerely hold this conviction leaves open the possibility that I might freely choose to depart from it. This possibility is due to the “frailty” (Rel, 29) of my will, that is, due to a propensity to knowingly depart from right reason in the face of temptation. Such awareness of the weakness of our will yields the most dramatic illustration of our rational imperfection.⁴⁸

VI. TYPES OF FREEDOM

I want to end by clarifying the conditions under which Kant endorses the view of free will I have expounded. Contemporary incompatibilists often

48. Although I cannot defend this point here in any detail, I want to suggest that Kant’s notion of a radical gap between reason- and desire-based motivation yields a more compelling account of desiring and acting against reason than ‘intellectualist’ views of desire ac-

hold that deliberation per se commits us to the falsity of determinism.⁴⁹ By contrast, Kant's motivation for incompatibilism, as I have reconstructed it, concerns only deliberation under practical laws. Why does Kant not appeal to the more general intuition that deliberation as such presupposes that determinism is false?

As we saw in Sections II and III, Kant accepts that determinism threatens neither the causal efficacy nor the epistemic openness of practical deliberation. He thus accepts the idea (often stressed by compatibilists) that even if determinism is true, how we act depends on our practical deliberation, and we cannot know what we will choose before we actually make our decisions.⁵⁰ For Kant, the presupposition that different courses of action are open to us in a 'thicker' metaphysical sense (i.e., in the sense that it not inevitable that we pursue one course rather than the other) derives entirely from the idea that two irreducibly different poles of motivation jointly constitute our nature as rational creatures: this accounts for the contrast between reason-based and desire-based motives that pervades our reflective self-awareness. For Kant, the existence of this contrast is guaranteed only by the antagonism between the moral reasons provided by practical laws and the desires that spring from our sensible nature: the nonmoral reasons of finite agents (and the subjective prudential norms that represent those reasons) themselves depend on the contingent desires whose satisfaction gives us happiness (KpV, 25–27). Thus, subjective norms are not categorically opposed to the ends of our contingent desires, and so our compliance with such norms is not essentially bound up with the presence of options to go against reason.⁵¹

cording to which the conflict between desire and reason can only be a matter of degree because desiring as such involves seeing a reason. For a classic formulation of the intellectualist view, see Thomas Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 2000), 20–40. For a forceful defense of Kantian anti-intellectualism about desiring, see Tamar Schapiro, "The Nature of Inclination," *Ethics* 119 (2009): 229–56. Kant's anti-intellectualism allows us to do justice to the phenomenology of feeling tempted to act and of acting in ways that one deems wholly contrary to reason: in the above example, I recognize that my egocentric desire to escape shame cannot be rationally weighed against the objective value of persons who would be the victim of my lie (GMS, 434–35). Notice that although my desire does not represent a reason, it is not an unintelligible 'urge' devoid of intentional content. Its motivational grip on me is due to its representation of the 'disagreeableness' that I associate with feeling ashamed. For an illuminating account of weakness of will within a Kantian framework, see Thomas Hill, *Virtue, Rules, and Justice* (Oxford: Oxford University Press, 2012), 107–59.

49. For a classic defense of this view, see Peter van Inwagen, *An Essay on Free Will* (Oxford: Clarendon, 1983).

50. This is the core of Bok's compatibilist response to van Inwagen. See Bok, *Freedom and Responsibility*, 96–114.

51. To confirm this, we can note that on a subjectivist view of reasons such as the one proposed by Bernard Williams, the occurrence of temptation or weakness of will is entirely contingent: the "subjective motivational set" of an agent may be such that she has a few

True, in the actual world we need the help of practical reason to ensure that we act prudently (KpV, 61). But it is (even empirically) possible that our recognition of the means toward our happiness might infallibly produce an unflinching motive to take those means. In this case, our nonmoral choices would need no assistance from practical reason, the discovery of means-end relations being a task for theoretical reason (KU, 172–73); we would never face a temptation to make imprudent choices, and thus rules of prudence would not strike us in the shape of oughts. By contrast, no contingent empirical desire could replace the impact that practical reason has on a choice that is made for necessary reasons; a finitely rational agent cannot fail to find herself with options to choose contrary to objective norms, and thus the objective norms of morality necessarily strike imperfectly rational beings as oughts.

Now imagine a world in which the empirical psychology of finite agents determines them to act in ways that efficiently satisfy their sensible needs. Under the further supposition that in this world no categorical imperatives exist so that morality is an illusion,⁵² Kant would accept compatibilism. If all our reasons for acting depend on our empirically given desires, then volition that is inevitably determined by correct representations of the means to our empirically given ends can be considered free because it answers to the only purpose we pursue in deliberating and acting.⁵³ This is why for Kant the awareness of a kind of freedom that is incompatible with determinism depends on our consciousness of the moral law whose validity is independent of contingent desires (KpV, 29–30).

Thus, Kant's conviction that alternative possibilities are essential to freedom concerns beings who satisfy two conditions: first, they are susceptible to a ubiquitous propensity to choose on behalf of their contingent desires; second, they are subject to norms that provide necessary reasons for acting. In the scenario just contemplated, where no objective norms exist, the second condition is violated: we lack a faculty of practical reason that is an autonomous source of ends in relation to which a choice based on a contingent desire could be deemed erroneous. In the case of divine agency, the first condition is violated: God's guidance by

dominant desires that never conflict and that she is always motivated to satisfy with maximal efficiency. See Bernard Williams, "Internal and External Reasons," in *Moral Discourse and Practice*, ed. Stephen Darwall, Allan Gibbard, and Peter Railton (Oxford: Oxford University Press, 1997), 363–72.

52. See GMS, 445. Kant's view that the nonexistence of (valid) categorical imperatives would entail that morality is an illusion is not idiosyncratic. For instance, John Mackie argues that because categorical imperatives do not exist, our ordinary moral judgments are false. See John Mackie, *Inventing Right and Wrong* (London: Penguin, 1991).

53. Kant imagines a similar scenario at GMS, 395. He grants that beings whose deliberation determines them to act in ways that satisfy their desires have the "psychological freedom" (KpV, 96) that is central to the empiricist tradition, e.g., to Hobbes or Hume.

reason is immune to the possibility of resisting reason because there is no metaphysical basis for such resistance. God has no sensible nature that afflicts him with options to go against reason.

My interpretation here suggests how Kant's view should be situated in the contemporary debate about the connection between norm-governed action and the possibility of violation. Kant, like Douglas Lavin, rejects Christine Korsgaard's view that norm-governed agency as such requires the possibility of violation.⁵⁴ For Kant, divine freedom consists solely in the ability to act rightly (Rel, 50). But Kant does not accept Lavin's or Susan Wolf's view that the freedom of finite agents is likewise identical to the ability to act for the right reasons.⁵⁵ For Kant, the justification of why divine agency is completely exempt from the possibility of error goes hand in hand with his justification of why free human action is afflicted with a radical, pervasive propensity to act wrongly: finite agents always choose under conditions of sensible affection by empirical desires that direct them away from the right reasons, whereas God must be conceived as being entirely unaffected by sensible conditions.

Kant here denies a presupposition shared by Lavin and Korsgaard, namely, the idea that we can give a unified account of what it takes to act freely under normative laws that applies indiscriminately to finite and holy agents.⁵⁶ For Kant, finite and holy agents exhibit different forms of freedom and agency because their respective wills are constituted differently: a Holy Will is simply identical with practical reason, whereas the human will comprises not just practical reason but also a sensibly affected faculty of choice (MS, 213, 222). The categorical difference between these two types of will is reflected in the forms of free agency that spring from them.

In all this, Kant's point is not that having the option to go against reason is a good thing. There is nothing desirable about falling short of practical perfection. The concept of a Holy Will represents a moral ideal which can (and should) inspire us toward moral self-improvement (Rel, 61). Alas, the practical perfection of the Holy Will cannot be anything

54. See Christine Korsgaard, "The Normativity of Instrumental Reason," in *Ethics and Practical Reason*, ed. Garrett Cullity and Berys Gaut (Oxford: Oxford University Press, 1997), 215–54, 240. Compare Korsgaard, *The Sources of Normativity*, 161. She suggests that we can only describe the acts of God. Her point seems to be that when we try to conceive a being that cannot fail to comply with normative rules, we lose our grip on the idea of normative guidance and must represent an automaton. But it is not clear why that should be (see Lavin, "Practical Reason and the Possibility of Error," 443–46). For Kant, we can draw a striking contrast here: an automaton is determined by nonrational mechanistic causes and is programmed to function by an external designer (KpV, 99), whereas God is not set up by external sources, and his agency is entirely unaffected by the nexus of nonrational causes.

55. See n. 31 above.

56. See Lavin, "Practical Reason and the Possibility of Error," 453; both he and Korsgaard address the question, "what is agency?" in the most general sense.

but an ideal for someone who finds herself in the human condition. For Kant, an adequate account of human free agency must do justice to this regrettable fact.

VII. CONCLUSION

I have offered a novel interpretation of Kant's incompatibilism. The key point of Kant's view, as I have reconstructed it, is that finite agency under normative laws presupposes that the agent is in a position to favor (in Kant's terms) either her sensible nature or her higher rational vocation: our ability to act on the basis of objective reasons has a shadow, namely, our propensity to choose against such reasons, that perpetually hovers over our acts of practical self-governance. This conception of the human condition accounts for the significance of real alternative possibilities in human agency. For Kant, such alternatives could not exist if determinism were true. I have tried to show that his view is rooted in an intriguing view of the metaphysics of normativity and in a resourceful moral psychology.