

# Kant on Free Will

*Derk Pereboom*

*Cornell University*

## *Abstract*

For Kant transcendental freedom consists in the power of agents to produce actions without being causally determined by antecedent conditions in exercising this power. He contends that we cannot establish whether we are actually or even possibly free in this sense. Kant claims only that our conception of ourselves as transcendently free involves no inconsistency, and that as a result the belief that we are free in this sense meets a relevant standard of minimal credibility. Justification of this belief ultimately depends on practical reasons: the need to believe that we are subject to moral obligation and that we are morally responsible. I argue that the belief that we are transcendently free does satisfy an appropriate standard of minimal credibility, but that the practical reasons Kant adduces for it should be controversial.

*Keywords:* Kant, Free will, Transcendental freedom, Determinism, Moral obligation, Moral responsibility.

## 1. Introduction

Immanuel Kant (1724–1804), dissatisfied with the sort of compatibilism advocated by philosophers such as David Hume (1711–1776), the hard determinism of Spinoza (1632–1677), and any view on which nature is indeterministic, advocated a bold conception of freedom on which human agents are the undetermined sources of their actions while at the same time all events in the natural world are thoroughly causally determined. In another respect, Kant’s theory is cautious: he maintains that our being free in this sense—transcendently free—cannot be established theoretically, that is, from evidence available to us. Rather, our evidence can establish only that believing we are transcendently free involves no inconsistency, and our justification for this belief must instead rely on practical reasons.<sup>1</sup>

<sup>1</sup> Spinoza presents his position on free will in his *Ethics* (Spinoza 1985), and Hume develops his view in the *Treatise of Human Nature* (Hume 1978) and in *An Enquiry Concerning Human Understanding* (Hume 2000). An earlier version of my statement of Kant’s view on free will can be found in Pereboom 2006.

## 2. The Terms of the Debate

It is best to begin with some definitions of key terms. The term ‘free will’ as it is used in the philosophical debate has several distinct senses, and the answer may depend on which sense is meant. In a first sense, to have free will is to have alternative possibilities for choice and action:

**Free will AP** (for “alternative possibilities”): free will is an agent’s ability, at a given time, either to act or to refrain: that is, if an agent acts with free will, then she instead could have refrained at that time from acting as she did.

A second sense, which predominates in the contemporary discussion, links free will to moral responsibility:

**Free will MR** (for “moral responsibility”): free will is an agent’s ability to exercise the control in acting required to be morally responsible for an action.

In the history of the free will debate, causal determinism of some sort has been taken to be the main threat to our having free will in each of these two senses:

**Causal determinism:** every event has causal antecedents that render it inevitable.

The parties to the free will debate are traditionally grouped into camps with reference to whether causal determinism and free will are compatible:

**Compatibilism:** our having free will is compatible with causal determinism, more specifically, with all of our actions being causally determined by factors beyond our control.

**Incompatibilism:** our having free will is not compatible with causal determinism, with all of our actions being causally determined by factors beyond our control.

Compatibilists in Kant’s philosophical ambit include, as mentioned, David Hume as well as G.W. Leibniz (1646–1716) and Christian Wolff (1679–1754). Incompatibilists divide into those who hold that determinism is false and that we have free will—the libertarians—and those who hold that determinism is true and that we lack free will—the hard determinists. In Kant’s context, the pietist philosopher and theologian Christian August Crusius (1715–1775) is an influential libertarian, and for him Spinoza is the main hard determinist opponent. For several key interlocutors, such as Gottfried Ephraim Lessing (1729–1781), it is unclear whether their sympathies ultimately lie with compatibilism or with Spinoza.

## 3. The Problem of Free Will as Kant Sees It

The outline of the problem Kant sets out for free will, beginning in the *Critique of Pure Reason* (1781, 2<sup>nd</sup> ed. 1787), is this: empirically—in the realm of appearance—which he also calls nature, every event, including each of our actions, is causally determined by temporally preceding conditions:

All the actions of the human being in appearance are determined in accord with the order of nature by his empirical character and other cooperating causes; and if we could investigate all the appearances of his power of choice down to their basis, then there would be no human action that we could not predict with certainty and recognize as necessary given its preceding conditions (A549–50/B577–78).

On Kant's proposal, empirical causal determinism does not rule out freely willed action. But he rejects any compatibilist resolution on which freely willed action is compatible with its causal history being *exhausted* by preceding natural conditions that causally determine its occurrence. In the *Critique of Practical Reason* (1788) Kant contends that such compatibilist freedom is ruled out because it cannot accommodate the control in action that freedom requires: "Since the past is no longer in my control, every action that I perform must be necessary by determining grounds *that are not within my control*, that is, I am never free at the point in time in which I act" (*KpV*, *Ak* V 94).

In this second *Critique* Kant also specifically rejects compatibilist views on which free actions are those that are determined by appropriate internal causes, as in Hume's view:

It is a wretched subterfuge to seek to evade this by saying that the *kind* of determining grounds of his causality in accordance with natural law agrees with a *comparative* concept of freedom... [in] which the determining natural cause is internal to the acting thing... And if the freedom of our will were the latter (say, psychological and comparative but not also transcendental, i.e. absolute), then it would at bottom be nothing better than the freedom of a turnspit, which, when once it is wound up, also accomplishes its movements of itself (*KpV*, *Ak* V 96–7).

If an action's causal genesis were exhausted by preceding natural conditions that causally determine its occurrence, the action would not be in the subject's power in a sense sufficient for what Kant calls *practical freedom*, which

presupposes that ... [an action's] cause in appearance was thus not so determining that there is not a causality in our power of choice such that, independently of those natural causes and even opposed to their power and influence, it might produce something determined in the temporal order in accord with empirical laws, and hence begin a series of occurrences *entirely from itself* (*ganz von selbst*) (A534/B562).

Practical freedom presupposes a power to choose independently of the natural causes that determine our actions. More exactly, practical freedom, in its negative aspect, is the power to produce an action without being causally determined by sensuous impulses (A534/B562), and in its positive aspect, which Kant also calls autonomy, it is the power to act motivated by moral principles whose source is rationality (e.g., *KpV*, *Ak* V 129). As Jerry Schneewind states it,

Kantian autonomy presupposes that we are rational agents whose transcendental freedom takes us out of the domain of natural causation. It belongs to every individual, in the state of nature as well as in society (A810/B838). Through it each person has a compass that enables "common human reason" to tell what is consistent with duty and what inconsistent (Schneewind 1998: 516).

Like Hume, Kant rejects the view that freely willed actions might simply be indeterministically caused or else uncaused events in nature, as in the view of Lucretius—such events would in his view amount to "blind chance" (*KpV*, *Ak* V

95).<sup>2</sup> Instead, the sort of indeterministic causation needed is “the power of beginning a state of *itself* (*von selbst*)—the causality of which does not in turn stand under another cause determining it in time in accordance with the law of nature” (A533/B561). The agent who acts freely *of itself* begins an effect in the sensible world, and, in doing so, is not determined by preceding causes (A541/B569). Kant calls this characteristic *transcendental* freedom, and he thinks of the power of transcendental freedom as an *intelligible cause*, by contrast with an empirical cause (e.g. A537/B565).<sup>3</sup>

An additional element in Kant’s theory of freedom is *Willkür*, “a power to do or to refrain from doing as one pleases (*ein Vermögen nach Belieben zu thun oder zu lassen*,” (*Metaphysics of Morals*, Ak VI 213), which he believes is necessary for moral obligation (to be discussed). But it is notable that for Kant, transcendental freedom need not involve having alternative possibilities for action, and it therefore does not entail having the power of *Willkür*. The following passage from the *Religion Within the Boundaries of Mere Reason* is indicative of Kant’s position on this issue:

There is no difficulty in reconciling the concept of *freedom* with the idea of God as a necessary being, for freedom does not consist in the contingency of an action (in its not being determined through any ground at all) i.e. not indeterminism ([the thesis] that God must be equally capable of doing good or evil, if his action is to be called free) but in absolute spontaneity. The latter is at risk only with predeterminism, where the determining ground of an action lies in antecedent time, so that the action is no longer in *my* power but in the hands of nature, which determines me irresistibly; since in God no temporal sequence is thinkable, this difficulty has no place (*Rel*, Ak VI 50n).

God cannot do otherwise than what is best but is yet free by virtue of divine action being produced entirely from the self (*ganz von selbst*) and thus not causally determined by conditions that precede it. For human beings, who have the power of *Willkür*, transcendental freedom may typically involve that power. But *Willkür*, as a matter of conceptual fact, is not required for transcendental freedom.

#### 4. Avoiding Hard Determinism

But now, given Kant’s empirical determinism about human action, and his indeterministic conception of free action, how can he avoid the Spinozistic result that because all actions are causally determined they are not free? Kant’s solution invokes his transcendental idealism, which distinguishes between things as they appear and things as they are in themselves, independently of how they appear.<sup>4</sup>

<sup>2</sup> Here Kant writes: “If, then, one wants to attribute freedom to a being whose existence is determined in time, one cannot, so far at least, except this being from the law of natural necessity as to all events in its existence and consequently as to its actions as well; for, that would be tantamount to handing it over to blind chance” (*KpV*, Ak V 95). Lucretius’s position is found in *De Rerum Natura*: “but what keeps the mind itself from having necessity within it in all actions [...] is the minute swerving of the first beginnings at no fixed place and at no fixed time” (Lucretius 1982: 2, 289-293).

<sup>3</sup> In the contemporary debate, agent-causal libertarians advance a position relevantly similar to Kant’s, e.g. Randolph Clarke (1993) and Timothy O’Connor (2000).

<sup>4</sup> This component of Kant’s position is set out by Allen Wood (1984: 83–89) and by Henry Allison (1990: 30–41).

While as appearances our actions are always causally determined by preceding temporal conditions, they may also originate from the self as it is in itself as their transcendently free cause. The core features of this solution are the following. First, human agents, as appearances, have an *empirical character*. Character, Kant specifies, is a law of something's causality, where causality is the activity of a cause (A539/B567). A thing's character is the way it behaves causally. A feature of how we as appearances, as empirical selves, behave causally is that our actions are causally determined by preceding natural conditions. However, the empirical character of our actions is compatible with our actions being produced by virtue of a character of a distinct kind, an *intelligible character* (A539/B567). In producing an action an agent, as a thing in itself, as an intelligible subject, may begin a state from the self without such causal determination.

In the *Critique of Pure Reason* Kant is careful to specify that we cannot *establish* that we as intelligible subjects are transcendently free; “we have not been trying to establish the reality of freedom”, or even to show that it is *really possible* i.e., metaphysically possible, that we are transcendently free; “we have not even tried to prove the possibility of freedom”. Rather, Kant aims only to ascertain that “nature at least *does not conflict with* causality through freedom” (A558/B586). That is, to ascertain that our conception of ourselves as transcendently free while embedded in the natural world is *not logically impossible*, i.e., that it is *logically possible*. For a conception to be logically possible is for it not to feature a contradiction (A244/B302). In Kant's view, a conception can be logically possible without its being a conception of a *real* possibility, because there are ways to preclude real possibility that do not involve establishing a contradiction. This perspective is in this respect similar to Saul Kripke's (1980) widely endorsed position according to which, for example, there is no contradiction involved in the conception that water is an element, while water's being an element, and thus not a compound, is nonetheless metaphysically impossible. Kant maintains that *we* can know (*erkennen*) real possibilities theoretically only through experience, that is, empirically, and in particular only by way of sensible intuition (A602/B630; cf. A218/B265–A226/B274), and this is akin to the contemporary view that we can know that water is a compound is metaphysically possible and that water is an element is metaphysically impossible only through experience. Accordingly, although our conception of ourselves as transcendently free features no contradiction, this is not enough to establish that this conception is metaphysically possible.

Kant thus denies that we can know that we, as we are in ourselves, have the power of transcendental freedom. I've argued that Kant's more general denial of knowledge of things in themselves, a core element of his transcendental idealism, amounts to rejecting what I call *substance-knowledge* of them (see Pereboom 1991). Substances in the Leibnizian scheme are things in themselves (and so not appearances) whose essential features include fundamental causal powers, and these powers are conceived as intrinsic properties of these entities. Specifically, for Leibniz, all substances are monads, modeled on Descartes's immaterial souls, whose essential feature is the fundamental causal power of representation, and the monad has this power intrinsically. In denying knowledge of things in themselves, Kant is ruling out knowledge of such fundamental intrinsic causal powers of substances. Transcendental freedom would indeed be such a fundamental causal power—an intrinsic feature of human selves as they are in themselves. Accordingly, knowledge of this power is ruled out by Kant's stricture. In accord with this account, although we can form a conception of transcendental freedom, we

lack the ability to investigate the nature of fundamental causal powers of the self to establish whether transcendental freedom is a capacity we actually have, or even whether this is metaphysically possible.

But what does Kant then mean when he says he has shown that nature does not conflict with transcendental freedom (A558/B586)? What if transcendental freedom is metaphysically impossible? I suggest that the best interpretation of Kant's claim is that there is no internal inconsistency in the conception of this power that we can formulate by our reason, and that there is no inconsistency between the claim that this description is true of us as noumenal agents and our best empirical theories about the natural world. This view allows that we could never comprehend its nature as a fundamental causal power.

### 5. Belief in Transcendental Freedom

That this is what Kant aims to establish regarding transcendental freedom fits with his main objective—to vindicate a “belief” that we are transcendently free. In his view, we cannot understand the fundamental causal powers of things in themselves, but we can nonetheless have legitimate beliefs about these powers. The notion of belief at issue is a “subjectively sufficient” but “objectively insufficient” conviction that is based on practical and not on theoretical—i.e. evidential—considerations (A822–3/B850–1).<sup>5</sup> In addition, Kant remarks that to believe in the practical sense is just to be guided practically by the content of the belief; this notion of belief “refers only to the guidance (*Leitung*) that an idea gives me, and the subjective influence on the advancement of my actions of reason that holds me fast to [the idea], even though I am not in a position to give an account of it from the speculative point of view” (A827=B855). A plausible suggestion that combines these ideas is that one has a belief in this practical sense when any reservations one might have about the truth of the proposition are set aside so that one's conviction in that proposition may guide one's thought and deliberation in acting.

Kant contends that for some such beliefs, although their evidential basis is inadequate for their justification, the practical benefits that derive from their capacity to guide our actions nevertheless justify our having them. Thus, for example, for these sorts of practical reasons one is justified in having a belief that God exists despite the absence of adequate evidential justification. At the same time, Kant indicates that in general such beliefs—at least when they are about things in themselves—should yet satisfy a theoretical requirement: they cannot be internally inconsistent or inconsistent with what we can establish theoretically. Why does he advocate this requirement? On the one hand, Kant maintains that the law of non-contradiction governs things in themselves. But still, if the practical rationality of belief is at issue, then not even inconsistency should in principle rule it out. Suppose someone with a cerebroscope credibly threatened to torture and kill me unless I were to believe some contradiction or other. Would it not then be practically rational for me to believe a contradiction (e.g., by taking a drug that would cause me to do so)? Or suppose that to believe that moral principles hold for us we would also have to believe some inconsistent proposition. Shouldn't Kant seriously consider that believing the inconsistency is then legitimate on practical grounds?

<sup>5</sup> For a thorough account of Kant's theory of belief, see Chignell 2007.

Perhaps Kant would seriously consider this proposal—he does not explicitly rule on the issue. But his concern is that if one is certain that a proposition is false, it won't be psychologically possible to believe it in order to secure some practical effect, and we can indeed be certain that contradictions are false. Kant discusses an example of a man lacking good moral sentiments, who, although he

might be separated from the moral interest by the absence of all good dispositions, yet even in this case there is enough left to make him fear a divine existence and a future. For to this end nothing more is required than that he at least cannot pretend to any *certainty* that there is *no* such being and *no* future life, which would have to be proved through reason alone and thus apodeictically, since he would have to establish them to be impossible, which certainly no rational human can undertake to do (A830/B858).

If the claims to the existence of God and a future life did feature inconsistencies, then their impossibility could be established. We could then be certain that these claims were false, and if we were, it would be psychologically impossible to have the conviction in them required to secure the desired practical effect. To avoid the certainty of the falsehood of these claims, it would have to be shown that they did not feature inconsistencies.

## 6. Is the Claim that We Are Transcendental Free Credible?

An objection to Kant's position is that propositions can lack credibility in the sense at issue for reasons other than overt inconsistency. For instance, a proposition's being judged very highly improbable might render it at most insignificantly more credible in this sense than an inconsistent proposition. Consider the proposition that intelligent beings from another planet will conquer the earth within the next year and subject all human beings to slavery, which for me is highly improbable. This proposition features no inconsistencies, but I nevertheless I cannot believe it in the sense at issue, i.e., I cannot set aside my reservations about its truth so that my conviction in it can guide my thought and deliberation about acting.

So even if the claim that we are transcendently free does not feature an overt contradiction, is it relevantly more credible than a proposition that does? One issue that bears on this question is whether the empirical subject that is causally determined to cause the action is to be understood as *identical* to the noumenal subject that is transcendently free and thus not causally determined to cause that action. In the *Critique of Practical Reason* Kant explicitly asserts this identity:

if one still wants to save [freedom], no other path remains than to ascribe the existence of a thing so far as it is determinable in time, and so too its causality in accordance with the law of *natural necessity, only to appearance, and to ascribe freedom to the same being as a thing in itself* (*KpV*, *Ak* V 95; cf. *Ak* V 97).

Many interpreters favor a *one-world* view on the relationship between appearances and things in themselves, according to which in general an appearance is numerically identical to a thing in itself or to a plurality of such things (e.g., Allison 1990). There are several key texts that lend support to the one-world view, but there are also those that might be construed to favor a *two-world* reading, according to which every appearance is numerically distinct from any thing in itself (e.g. Jauernig 2012). The concern that arises on the one-world conception is that the

empirical self and the self as it is in itself would be identical while having incompatible properties. An empirical self is, in Kant's view, a complex of psychological states that can potentially be apprehended by inner sense. By his account, for any action it performs in the empirical world, the empirical self is causally determined to perform that action by preceding natural conditions. However, we are asked to believe that when the action is free, the self as it is in itself produces the action from itself without being causally determined by preceding conditions. Could the empirical self (E) and the self as it is in itself, also known as the noumenal self (N), be identical? (E) and (N) differ in their properties: (E) has the property of being causally determined by conditions beyond her control to cause the action, while (N) lacks this property by virtue of causing the action from herself. How is it at all credible that (E) and (N) should be identical while differing in this specific way?

This problem can be addressed by the two-world reading, since then E and N would not be identical. But on a two-world reading, the action, as an event in the empirical world, would be overdetermined in a peculiar way. By one strand of its causal history the empirical action has a sufficient cause in a transcendently free self, while by another strand this same action has a sufficient cause in a natural sequence that traces back to a time before the agent came to exist. Now there would be nothing incredible about the proposal that a transcendently free agent should make a free choice *on some particular occasion* for an action that was at the same time causally determined by a natural causal sequence. However, a more substantial proposal is required, which fares differently. It is that *all* transcendently free choices should be for just those actions that are at the same time determined to occur by virtue of natural causal sequences, and that *none* of these choices be for alternatives to those actions. If we were selves as they are in themselves making transcendently free choices for our actions, it seems that one would *expect*, in the long run, that these choices would be manifest in the empirical world as patterns of divergence from the deterministic natural laws. The proposal that there are no such divergences, although it involves no contradiction, would appear to run so sharply counter to what we would expect to occur as to render it incredible.<sup>6</sup>

Perhaps a further story can be told to make this two-world proposal credible. According to the position of Luis de Molina on divine providence, God knows, eternally, what every possible libertarian free creature would choose in every possible circumstance, and with this knowledge, God is able to direct the course of history with precision, partly in virtue of creating just those free creatures whose choices fit a preconceived divine plan.<sup>7</sup> On one version of this Molinist view adapted to Kant's idealism, God would reconcile noumenal transcendental freedom with phenomenal determinism by creating just those transcendently free beings the appearances of whose free choices conform to the deterministic laws that God intends for the phenomenal world. Now Molinism is actually endorsed by many of those who have thought seriously about free will and divine providence, and this Kantian version of Molinism might not be significantly less credible. So there might indeed be some who would be able to set aside any reservations they might have about this view, with the result that the conviction that we are

<sup>6</sup> See Pereboom 2001: 79–85 for further development of this argument.

<sup>7</sup> Molina 1988. For a thorough exposition and defense of Molina's position, see Flint 1998. See also Christopher Insole's (2013) discussion, especially Chapter 9.

transcendentally free can provide guidance for their actions. But for others this Kantian version of Molinism may not be credible due to the idealistic or the theistic beliefs that it presupposes. Or else, it may not be credible because of the main objection to Molinism, that there could be no truths about what possible transcendentally free creatures would choose not-yet actual or counterfactual circumstances, since any candidate for a ground for such truths would actually only ground truths of the form: *S would probably choose A*, and *not S would in fact choose A*.<sup>8</sup>

## 7. Transcendental Freedom and Moral Obligation

Even though according to Kant it is far from true that we can establish on the basis of the evidence that we are transcendentally free, he nevertheless contends it is legitimate for us to believe that we are. The grounds for legitimacy are practical—we have reasons that derive from morality in particular for believing that we are transcendentally free.<sup>9</sup> Kant presents two such reasons. The first is that moral imperatives have the form of ‘ought’ judgments, and the truth of such judgments is incompatible with the determinism we find in nature, where each event is causally determined by preceding conditions. The second is that our judgments of moral responsibility are incompatible with determinism, or again at least with this sort of natural determinism.

On ‘ought’ judgments Kant claims:

Now that this reason has causality, or that we can at least represent something of the sort in it, is clear from the *imperatives* that we propose as rules to our powers of execution in everything practical. The *ought* expresses a species of necessity and a connection with grounds which does not occur anywhere else in the whole of nature. In nature the understanding can cognize only *what exists*, or has been, or will be. It is impossible that something in it *ought to be* other than what, in all these time-relations, it in fact is; indeed the ought, if one merely has the course of nature before one’s eyes, has no significance whatever (A547/B575).

But, Kant claims, “perhaps everything that *has happened* in the course of nature, and on empirical grounds inevitably had to happen, nevertheless *ought not to have happened*” (A550/B579). Or at least this is a moral claim we would assume to be true.

In the *Religion Within the Boundaries of Mere Reason*, for example, Kant explicitly endorses the “ought implies can” principle; “For from the practical point of view this idea [of a prototype of humanity pleasing to God] has complete reality within itself. For it resides in our morally-legislative reason. We *ought* to conform to it, and therefore we must *be able* to” (*Rel, Ak VI*: 62).<sup>10</sup> The following moral “ought implies can” principle is indeed attractive: if one ought to do something, then it must be the case that one can do it.<sup>11</sup> Accordingly, if because in general

<sup>8</sup> A classic source of the “grounding” objection is Adams 1979.

<sup>9</sup> See Markus Kohl (2023) for an extensive discussion of these issues.

<sup>10</sup> In addition, Kant asserts “ought implies can” at *Rel, Ak VI*: 45: “For, in spite of that fall, the command that we *ought* to become better human beings still resounds unabated in our souls; consequently, we must also be capable of it”; and he defends a similar idea at *Rel, Ak VI*: 68: “Yet duty commands that he be good, and duty commands nothing but what we can do”.

<sup>11</sup> For a sensitive discussion of these issues, see Sinnott-Armstrong 1984. For discussions of Kant on obligation and ability, see Stern 2004 and Kohl 2015, 2023.

one is causally determined to act as one does, one can never do otherwise, it would be false that one ever ought to do otherwise. In particular, if whenever one acts wrongly, it is never true that one ought to do otherwise, what would be the point of a system of moral 'oughts'? It would seem that if "A ought to do  $x$ " is true at all, it must be true not only when A does  $x$ , but also when A fails to do  $x$ .<sup>12</sup> If all of this is sound, then given that we have a good practical reason to preserve moral 'ought' judgments, we have a good practical reason to believe that we are transcendentally free.

But how strong is this practical reason? An initial problem is that what would be required for moral principles to be true or hold for us is not the belief that we are transcendentally free, but rather we actually are transcendentally free. By contrast, sometimes a belief itself, and not specifically the truth of the content of the belief, is what is needed to secure a practical goal. For example, in the version of the moral argument for belief in God that we find in the *Religion*, Kant's idea is that without a belief in God we could not also believe that in each person happiness will eventually be proportioned to virtue, and the suggestion is that if we did not have this belief about happiness and virtue, we would be disheartened to the degree that our motivation to moral action would suffer (*Rel, Ak VI 6–8n*; see also Adams 1979). In this case, it is the *belief* that God exists, specifically, that would prevent the hindrance to moral motivation. But, returning to the case of freedom, perhaps it is the belief that moral principles are true, rather than the truth of the moral principles, that Kant aims to secure. Given this supposition, the belief that we are transcendentally free might indeed be what is required to secure the goal.

Two worries one might raise about this proposal are: first, perhaps, in our conception of morality, there are 'ought' judgments sufficient for morality that do not presuppose an "ought implies can" principle; and second, there may be principles sufficient for morality that are not 'ought' judgments and are not undermined by an "ought implies can" principle. So first, one clear role that moral 'ought' judgments have is to guide actions. We say to people that they ought not steal, for instance, in order guide their practical reasoning so that they might refrain from stealing. Moreover, for Kant, an important function of practical judgments generally is to guide actions. Does an "ought implies can" principle need to be true for 'ought' judgments to have this action-guiding function? Not obviously. Suppose that causal determinism is true, and that hence no agent could ever do otherwise. Nevertheless, we do not typically know in advance—before deliberation is complete or a decision has been made—which choice for action has been causally determined. Rather, it is almost always true that from the epistemic point of view of the agent at the time of deliberation more than one option for which choice she will make is possible. That is, more than one such option is possible relative to an appropriate subset of what she believes (or of what she should believe), in the sense that it is at least not ruled out by what she believes (or by what she should believe).<sup>13</sup> Often, when one attempts to guide an agent by means of a moral 'ought' judgment, it is the range of options for action that are in this sense epistemically possible for the agent at the time of deliberation that one addresses. Frequently, it is significantly probable that expressing a moral 'ought' judgment will causally influence action, and thus there is a good moral

<sup>12</sup> See Haji 2002 for a more thoroughly developed argument for this claim.

<sup>13</sup> See Kapitan 1986.

reason to do so—even if it turns out that because causal determinism is true the agent could not have complied with the judgment.

Against this solution one might argue that although ‘ought’ judgments in these action-guiding roles would retain practical value, so that it might often be practically rational to express them, they must nevertheless be false if causal determinism were true. But even Ishtiyaque who has argued at length that ‘ought’ judgments as instruments for deontic appraisal of actions would be false if determinism were true, agrees that this type of undermining argument does not hold for ‘ought’ judgments when they have an action-guiding function.<sup>14</sup> Haji presupposes, as C.D. Broad also contended (see Broad 1952), that ‘ought’ judgments have various distinct roles, and that these roles have different truth or assertability conditions. To my mind he is right to suggest that the action-guiding variety of ‘ought’ judgment can be retained even if determinism is true and no agent could have done otherwise, and even if this consideration undermines ‘ought’ judgments in other roles.

Secondly, even if moral ‘ought’ judgments do turn out not true or do not hold for us because our actions are causally determined, we need not also accept that no moral principles are true or hold for us. For, plausibly, moral judgments about rightness and wrongness of actions could still be true. Suppose that someone is causally determined by genetic predisposition and childhood abuse to be a violent criminal. His actions are, intuitively, still morally wrong, and it is still morally wrong for him to commit these crimes. Moreover, moral judgments such as “it is morally good for A to do *x*” and “it is morally bad for A to do *y*” still could be true. Thus, for example, even if one is causally determined to refrain from giving to charity, and even if it is therefore false that one ought to give to charity, it still might still be good to do. Embezzling funds from one’s company would be a bad thing to do, even if one’s act is causally determined, and hence, even if it is false that one ought not to do so. It would seem that principles regarding moral rightness and wrongness, goodness and badness, can fulfil most or even all of the roles that moral ‘ought’ judgments have in guiding action, in, for example, moral encouragement, admonition, and evaluation. If this is so, and since the truth of these alternative principles does not appear to require transcendental freedom, the practical reason for believing that we are transcendently free that Kant adduces at this point would carry insufficient weight.

## 8. Transcendental Freedom and Blameworthiness

But even if the consciousness that there is a moral law does not all by itself support a belief in transcendental freedom, another aspect of our sense of morality does. When people perform actions contrary to the moral law, many of us typically judge them blameworthy in a sense that features basically deserved pain or harm. More specifically,

For an agent to be *morally responsible for an action in the basic desert sense* is for the action to be attributable to her in such a way that if she was sensitive to its being morally wrong, she would deserve to be blamed or punished in a way that would be harmful to her, and if she was sensitive to its being morally exemplary, she would deserve to

<sup>14</sup> Haji (2002: 77) writes: “For the argument for the incompatibility of determinism and deontic morality is not in any way concerned with the action-guiding function of ought judgments”.

be praised or rewarded in a way that she would be beneficial to her. The desert at issue is basic in the sense that the agent, to be morally responsible, would deserve such blame or punishment, praise or reward, just by virtue of having performed the action with sensitivity to its moral status, and not, for example, by virtue of consequentialist or contractualist considerations (Pereboom 2021: 11–12).<sup>15</sup>

Basic desert moral responsibility contrasts with the sort that appeals only to non-basic desert, which invokes further goods, such as good consequences, to justify desert claims. It also contrasts with conceptions of moral responsibility that do not invoke desert at all, but are instead resolutely forward-looking, recruiting blame only to advance goods such as moral formation and restoration of relationships impaired by wrongdoing.

For many, the intuition that wrongdoers deserve to be punished concerns basic desert specifically, in contrast with its non-basic relative. In *The Metaphysics of Morals* Kant provides an example that illustrates this sense, one that features a murderer in an imagined island society that is about to dissolve itself, in which there is no further good to which punishing a wrongdoer would contribute. Kant strenuously advocates that he should be executed, just because of the crime he has committed; that is, for reasons of retributive desert alone (*Ak* VI: 331–33). To embellish the example, imagine a person on an isolated island viciously murders everyone else on the island and that he is not capable of moral reform due to ingrained hatred and rage. Thus, there are no good consequences that the punishment might aim to realize, and there is no longer a society on the island whose rules might be determined by contract. Many nonetheless have the intuition that this murderer deserves to be punished severely. The desert would be basic since the specifics of the example eliminate non-basic desert.

Accordingly, the second practical consideration Kant adduces in support of the belief that we are transcendently free is that if we lacked this kind of freedom our judgments of blameworthiness—and moral responsibility more generally—would turn out to be false. The issue is discussed in the “malicious lie” passage in the *Critique of Pure Reason*:

one may take a voluntary action, e.g. a malicious lie, through which a person has brought about a certain confusion in society; and one may first investigate its moving causes, through which it arose, judging on that basis how the lie and its consequences could be imputed to the person. With this first intent one goes into the sources of the person’s empirical character, seeking them in a bad upbringing, bad company, also finding them in the wickedness of a natural temper insensitive to shame, partly in carelessness and thoughtlessness; in doing so one does not leave out of account the occasioning causes. In all this, one proceeds as with any investigation in the series of determining causes for a given natural effect. Now even if one believes the action to be determined by these causes, one nonetheless blames the agent, and not on account of his unhappy natural temper, not on account of the circumstances influencing him, nor even on account of the life he has led previously; for one presupposes that it can be entirely set aside how that life was constituted, and that the series of conditions that transpired might not have been, but rather that this deed could be regarded as entirely unconditioned in regard to the

<sup>15</sup> This is a slight revision relative to the formulation in Pereboom 2021; 11–12. For earlier formulations, see Pereboom 2001: xx, and 2014: 2.

preceding state, as though with that act the agent had started a series of consequences entirely from himself. (A554–55/B582–83).<sup>16</sup>

The idea is that we have good practical reason to judge the liar blameworthy, and since blameworthiness requires transcendental freedom, we thereby have a good practical reason to believe that he is transcendently free.

But note the epistemic situation that Kant thinks we are in. We cannot show on the basis of the evidence that we are transcendently free, or even that transcendental freedom is causally possible, but only that a description of transcendental freedom is not internally inconsistent or inconsistent with our best theories about the empirical world. Given this epistemic situation, and assuming Kant's incompatibilism, would it be morally acceptable to judge a wrongdoer blameworthy for what he has done? Or to justify expressing one's anger towards him by the claim that he is blameworthy? Or, if he is a criminal, to deprive him of his liberty or life on the ground that he deserves such treatment just by virtue of having done wrong?

Consider, for example, the murderer in the imagined island society that is about to dissolve itself, who Kant believes basically deserves to be executed. Imagine that the murderer protests that he was determined by empirical causes to act as he did. Would the following reply be morally acceptable? "Although we have no evidence that you are transcendently free, and although we cannot establish that such a power of agency is even metaphysically possible, yet our belief that you are free in this way involves no inconsistency, and we need to have this belief in order to justify treating people like you as blameworthy and basically deserving of punishment". This is dubious. Treating an offender as basically deserving of the harm or pain involved in blaming and punishing him requires a high standard of justification—much higher than the standard of consistency Kant endorses. If one's justification for the harm or pain depended on the claim that we are transcendently free, and we have little or no evidence for this attribution, and the story we need to tell to reconcile transcendental freedom with our best empirical theories at best marginally credible, then that justification would appear to be inadequate.

In addition, if we relinquish basic desert, we yet have access to a model for holding moral responsible that arguably can serve the same functional role but does not presuppose transcendental freedom. Rather than confronting wrongdoing by blame suffused with retributive sentiments, we can turn to a forward-looking stance of moral protest instead. Pamela Hieronymi has proposed that blame should be understood as moral protest (see Hieronymi 2001: Chpt. 2). But while in her view the negative reactive attitudes have an essential role in blaming as protest, we can set this part of her account aside. Moral protest can be viewed as a psychological stance, a posture of mind that has certain aims or functions that are manifest as dispositions to act. It is a stance of opposition to specific immoral actions and their

<sup>16</sup> The "malicious lie" passage continues: "This blame is grounded on the law of reason, which regards reason as a cause that, regardless of all the empirical conditions just named, could have and ought to have determined the conduct of the person to be other than it is. And indeed one regards the causality of reason not as a mere concurrence with other causes, but as complete in itself, even if sensuous incentives were not for it but indeed entirely against it; the action is ascribed to the agent's intelligible character: now, in the moment when he lies, it is entirely his fault; hence reason, regardless of all empirical conditions of the deed, is fully free, and this deed is to be attributed entirely to its failure to act" (A555/B586).

general type, whose aims include communication of this opposition to wrongdoing, together with reasons to refrain from it. Such a model can specify that the goals of blaming are the forward-looking aims of moral formation of wrongdoers, protection against wrongdoing, and reconciliation in relationships impaired by wrongdoing.

This model can also accept the Kantian proviso that it is the agent's faculty of recognition and responsiveness to reasons that is engaged in moral protest. In that process we may request an explanation with the intent of having the agent acknowledge a motivation to act wrongly, and then, if he has in fact so acted without excuse or justification, we may intend for him to come to see that the motivation issuing in the action is best eliminated. This change is produced by way of the agent's recognition of moral reasons to make it. More generally, it is a wrongdoer's responsiveness to reasons together with the forward-looking aims that explains why he is appropriately addressed by moral protest so characterized. This model of holding morally is not tied to transcendental freedom because given its forward-looking aims, it is compatible with causal determinism.

## 9. Transcendental Freedom as a Fact of Reason

Kant pursues what may appear to qualify as a more ambitious route to affirming transcendental freedom in the *Critique of Practical Reason*. In this the second *Critique* Kant argues that the reality of our having freedom in this sense is established as a *fact of reason*—*Faktum der Vernunft* (*KpV* 5: 47), and he specifies that we have a practical cognition (*praktischen Erkenntnis*) of this fact (*KpV* 5: 103). On Andy Reath's (2024) account, it is in proper exercises of our power of reason in moral judgment that the fact of reason is given to us, and this fact is *the reality of pure practical reason and the authority of the moral law*. The proper exercises of the power of reason are instances of moral judgments that assign deliberative priority to specific moral grounds, for example truthfulness over personal advantage. Such judgments affirm the Categorical Imperative, specifically the formula of universal law—act only on that maxim that you can at the same time will as a universal law—as their formal principle. This given fact of reason yields practical cognition of the authority of the fundamental principle of morality. The reality of free will is subsequently given through the fact of reason, also as a practical cognition. The moral judgments at issue are judgments of moral obligation, and because 'ought' implies 'can', they presuppose that in cases in which one has not done what one ought, one could have done otherwise.

The account of the *Critique of Pure Reason* differs in that it invokes neither the notion that freedom is given as or via a fact of reason, nor that it is a practical cognition. The account of the second *Critique* might at least initially be conceived as a daring new move, aiming to show, as Karl Ameriks (2000: 189–233) puts it, *that* we are transcendently free. But as Henry Allison (1990) points out, there are at least two reasons for restraint that derive from Kant's views more generally on any interpretation of the claim that the reality of free will is secured as or by a fact of reason. One is that Kant affirms, even in the second *Critique*, that there are no instances in which we can ascertain that we are motivated to act on the basis of the moral law, by contrast with self-interested incentives (*KpV* 5: 47). Accordingly, it is not in moral decision or action itself that the reality of practical reason and the authority of the moral law are given to us as a fact, but rather in judgments as to how the moral law enjoins us to decide and to act. In turn, our transcendental freedom is not revealed in our deciding or acting morally, but as a presupposition of the

authority of the moral law. When I recognize the authority of the moral law in cases in which I acted wrongly, I see that this recognition presupposes that I could have acted otherwise.

A second reason for restraint is that establishing the reality of practical reason and the authority of the moral law does not issue in a theoretical cognition, but rather in a practical cognition. But what exactly is a practical cognition? Kant specifies that, unlike a theoretical cognition, a practical cognition does not involve an intuition:

Consciousness of this fundamental law may be called a fact of reason because one cannot reason it out (*herausvernünftlen*) from antecedent data of reason, for example, from consciousness of freedom (since this is not antecedently given to us) and because it forces itself upon us of itself as a synthetic a priori proposition that is not based on any intuition, either pure or empirical (*KpV* 5: 31).<sup>17</sup>

As a consequence, such a practical cognition does not meet the standard of knowledge of synthetic judgments central to the critical project, i.e., that it is based on intuition.

Kant makes apparently stronger and weaker claims about the epistemic status of the fact of reason. For an apparently stronger claim, he writes in *Critique of Practical Reason* that in the first *Critique* “theoretical reason was forced to assume at least the possibility of freedom to fill a need of its own”, but now states that “the moral law proves its reality [...] by the concept of reason determining the will immediately” (*KpV* 5: 48). Yet when Kant comes to stating precisely what he means, he is circumspect. The concept of causality as it applies to freedom has reference only to the practical, and “beyond this they lay no claim to the cognition of [intelligences]”. Moreover,

as for whatever other properties belonging to the supersensible way of representing those things may be brought forward in connection with these categories, these are without exception to be counted not as knowledge [*Wissen*] but only as warrant [*Befugnis*] (for practical purposes, however, a necessity) to admit [*anzunehmen*] and presuppose [*voraussetzen*] them, even where supersensible beings (such as God) are assumed by analogy, that is by purely rational relation of which we make a practical use with respect to what is sensible; and so, by this application to the supersensible but only for practical purposes pure theoretical reason is not given the least encouragement to fly into the transcendent (*KpV* 5: 57).

Here is an interpretation that I find attractive. When we take the practical point of view in making moral judgments, the content and authority of the moral law are, from the perspective of this moral consciousness, presented to us as a fact. Implicit in this fact, and derivable from it, is the judgment that we are transcendentally free. But this does not provide theoretical knowledge that we are transcendentally free—this would require an intuition that is unavailable to us. From the theoretical point of view the fact of freedom given to us practically does not amount to knowledge (*Wissen*), but rather to a practically warranted presupposition.

<sup>17</sup> See also (*KpV* 5: 55): “The objective reality of a pure will or, what is the same thing, of a pure practical reason is given a priori in the moral law, as it were by a fact—for so we may call determination of the will that is unavoidable even though it does not rest on any empirical principles”.

On this interpretation, the accounts of the two *Critiques* are consistent. Both agree that we have no intuition and no theoretical knowledge of our transcendental freedom. But the *Critique of Practical Reason* adds that from the practical perspective, when we are making moral judgments, the moral law and its authority is presented to us as a fact, as is, derivatively, our transcendental freedom. Furthermore, we are practically justified in taking the authority of the moral law and human transcendental freedom as presuppositions in moral deliberation.

## 10. Final Words

Kant's theory of transcendental freedom stands out as a signature attempt to reconcile our views of ourselves as free and responsible agents with an increasingly sophisticated scientific picture of the world, and it has been an important philosophical inspiration and source for many who have this ambition. As a rule, determinists about nature are either hard determinists or compatibilists about free will. But these positions arguably relinquish widespread intuitions about the capacities for action we have, or about what is required for moral responsibility. Kant's theory is especially ambitious in that it aims to preserve these intuitions by developing a view of free will on which we, as substances, have the power to cause actions without being causally determined to cause them. But he maintains that this cannot be established or even shown to metaphysically possible on the basis of evidence we available to us. Instead, he claims only that we can show that our conception of our having free will in this sense involves no inconsistency, and that the legitimacy of a belief that we have this kind of freedom must rely on practical reasons.

To my mind, there are several important respects in which Kant's treatment of free will and action is highly plausible. Moral responsibility must indeed be grounded in transcendental freedom, but whether we have it cannot be established on evidence available to us. Yet Kant provides a consistent conception according to which we have free will of this sort. In addition, the practical reasons that Kant adduces for belief in transcendental freedom—that it is required for moral obligation and for blameworthiness—carry significant weight. However, Kant's justification for believing that we are transcendently free based on these reasons is subject to challenge. But there may be more to these reasons than I have seen, and there are other reasons for believing in transcendental freedom that may also need to be considered. Moreover, the credibility and commitments of alternative positions on free will must be weighed against Kant's. In consequence, the general contours of Kant's position on free will remain a significant and attractive option.

## References

### *Kant's works*

Citations of 'Ak' refer to Immanuel Kant, *Kant's gesammelte Schriften*, edited by the *Königliche Akademie der Wissenschaften* and its successors (Berlin: George Reimer (subsequently W. de Gruyter), 1902–).

A/B *Critique of pure reason*. English quotations are from the translation by Paul Guyer and Allen Wood, *The Cambridge edition of the works of Immanuel Kant: the Critique of pure reason* (Cambridge: Cambridge University Press, 1997). Any alterations are accompanied by the German equivalent.

- KpV Critique of practical reason*. English quotations are from the translation by Mary Gregor, *The Cambridge edition of the works of Immanuel Kant: practical philosophy* (Cambridge: Cambridge University Press, 1996), unless otherwise indicated.
- KU Critique of the power of judgment*. English quotations are from the translation by Paul Guyer and Eric Matthews, *The Cambridge edition of the works of Immanuel Kant: Critique of the power of judgment* (Cambridge: Cambridge University Press, 2000).
- Rel Religion within the boundaries of mere reason*. English quotations are from the translation by Allen Wood and George di Giovanni, *The Cambridge edition of the works of Immanuel Kant: religion and rational theology* (Cambridge: Cambridge University Press, 1996).
- VpR Lectures on philosophical theology*. English quotations are from the translation by Allen Wood and Gertrude Clark (Ithaca: Cornell University Press, 1978).
- Adams, R.M., 1977. Middle knowledge and the problem of evil. *American philosophical quarterly*, 14, 109–117.
- Adams, R., 1979. Moral arguments for theistic belief. In: C.F. Delaney, ed., *Rationality and religious belief*. Notre Dame: Notre Dame University Press, 116–140; repr. in Adams, R., 1987. *The virtue of faith*. Oxford: Oxford University Press, 144–163.
- Allison, H., 1990. *Kant's theory of freedom*. Cambridge: Cambridge University Press.
- Ameriks, K., 2000. *Kant's theory of mind, new edition*. Oxford: Oxford University Press.
- Broad, C.D., 1952. Determinism, indeterminism, libertarianism. In: Broad, C.D., *Ethics and the history of philosophy*. London: Routledge and Kegan Paul, 195–217.
- Chignell, A., 2007. Belief in Kant. *The philosophical review*, 116 (3), 323–360.
- Clarke, R., 1993. Toward a credible agent-causal account of free will. *Noûs*, 27, 191–203.
- Flint, T.P. 1998. *Divine Providence: the Molinist account*. Ithaca: Cornell University.
- Haji, I., 1998. *Moral appraisability*. New York: Oxford University Press.
- Haji, I., 2002. *Deontic morality and control*. Cambridge: Cambridge University Press.
- Hieronymi, P., 2001. Articulating an uncompromising forgiveness. *Philosophy and phenomenological research*, 62, 529–555.
- Hume, D., 1978. *A treatise of human nature*. Oxford: Oxford University Press.
- Hume, D., 2000. *An enquiry concerning human understanding*. Oxford: Oxford University Press.
- Insole, C.J., 2013. *Kant and the Creation of Freedom*. Oxford: Oxford University Press.
- Kapitan, T., 1986. Deliberation and the presumption of open alternatives. *The philosophical quarterly*, 36, 230–251.
- Kohl, M., 2015. Kant and 'ought implies can'. *The philosophical quarterly*, 65, 690–710.
- Kohl, M., 2023. *Kant on Freedom and Rational Agency*. Oxford: Oxford University Press.
- Kripke, S., 1980. *Naming and necessity*. Cambridge, MA: Harvard University Press.
- Jauernig, Anja. 2021. *The World According to Kant*. Oxford: Oxford University Press.
- Lucretius, 1982. *De rerum natura*. Translation by W.H.D. Rouse. Loeb Classical Library, Cambridge: Harvard University Press.
- Molina, L. de., 1988 [1595], *Liberi arbitrii cum gratiae donis, divina praescientia, providentia, praedestinatione et reprobatione*. Translation (of Part IV) by A.J., Freddoso. *On divine foreknowledge: part IV of the Concordia*. Ithaca: Cornell University Press.
- O'Connor, T., 2000. *Persons and causes*. New York: Oxford University Press.

- Pereboom, D., 1991. Is Kant's transcendental philosophy inconsistent? *History of philosophy quarterly*, 8, 357–372.
- Pereboom, D., 2001. *Living without free will*. Cambridge: Cambridge University Press.
- Pereboom, D., 2006. Kant on transcendental freedom. *Philosophy and phenomenological research*, 73 (3), 537–567.
- Pereboom, D., 2014. *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.
- Pereboom, D., 2021 *Wrongdoing and the moral emotions*. Oxford: Oxford University Press.
- Reath, A., 2024. Rational powers and the “fact of reason”. In: W. Gobsch and T. Land, eds., *The Aristotelian Kant*, Cambridge: Cambridge University Press, 180–206.
- Schneewind, J.B., 1998. *The invention of autonomy*. Cambridge: Cambridge University Press.
- Sinnott-Armstrong, W., 1984. ‘Ought’ conversationally implies ‘can’. *The Philosophical review*, XCIII, 249–261.
- Spinoza, B., 1985. *The collected works of Spinoza*. E. Curley, ed. and tr., Vol. 1. Princeton: Princeton University Press.
- Stern, R. Does ‘ought’ imply ‘can’? And did Kant think it does? *Utilitas*, 16 (1), 42–61.
- Watkins, E., 2005. *Kant and the metaphysics of causality*. Cambridge: Cambridge University Press.
- Wood, A.W., 1984. Kant's compatibilism. In: A.W. Wood, ed., *Self and nature in Kant's philosophy*. Ithaca: Cornell University Press, 73–101.