

# Kant on Transcendental Freedom

Derk Pereboom

*Philosophy and Phenomenological Research* 73, 2006, pp. 537-67.<sup>1</sup>

Kant is remarkable for his attempt to reconcile an essentially libertarian view of freedom and moral responsibility with a deterministic conception of nature.<sup>2</sup> As a rule, and as one might

---

<sup>1</sup> 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.

G *Groundwork of the Metaphysic of Morals*. English quotations are from the translation by H. J. Paton, *Groundwork of the Metaphysic of Morals* (New York: Harper and Row, 1964).

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).

<sup>2</sup> Or at least I shall argue that this is what Kant aims to do. On this characterization of Kant's project, I follow Henry Allison in *Kant's Theory of Freedom* (Cambridge: Cambridge University Press, 1990), e.g. pp. 1-2, 29-46, but I disagree sharply with Hud Hudson's reading in *Kant's Compatibilism* (Ithaca: Cornell University Press, 1994), which is similar in certain

expect, determinists about nature are either hard determinists or compatibilists about freedom. But these positions arguably give up 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 freedom akin to agent-causal libertarianism, while at the same time accepting an uncompromising scientific determinism about the natural world. In another respect, Kant's theory is not ambitious: he maintains that it cannot be established theoretically -- i.e., on the basis of any evidence available to us -- that we have this sort of freedom, or even that it is metaphysically possible that we do. Rather, he claims only that our conception of our being free 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.<sup>3</sup>

---

respects to Ralf Meerbote's reconstruction, for example in "Kant on the Nondeterminate Character of Human Actions," in *Kant on Causality, Freedom, and Objectivity*, ed. W. L Harper and Ralf Meerbote (Minneapolis: University of Minnesota Press, 1984), pp. 138-63, and "Kant on Freedom and the Rational and Morally Good Will," in *Self and Nature in Kant's Philosophy* (Ithaca: Cornell University Press, 1984), pp. 57-72. I also disagree, but less extensively, with Allen Wood's compatibilist interpretation in "Kant's Compatibilism," in *Self and Nature in Kant's Philosophy* (Ithaca: Cornell University Press, 1984), pp. 73-101. These differences will become evident in what follows.

<sup>3</sup> This is, at any rate, Kant's critical position on transcendental freedom, the best expression of which is found in the *Critique of Pure Reason*, A532/B561-A558/B586. As Karl Ameriks points out, at various points in his career Kant maintained that he could show *that* we are transcendently free (*Kant's Theory of Mind, New Edition* (Oxford: Oxford University Press, 2000), pp. 189-233; Ameriks's historical discussion here is very instructive). Even after the publication of the second edition of the *Critique of Pure Reason* (1787) – in the *Critique of Practical Reason* (1788) – Kant claims that through our consciousness of the moral law we *cognize* (*erkennen*) transcendental freedom, albeit from a practical point of view:

Therefore, that unconditioned causality and its faculty, freedom, and therewith a being (myself) which belongs to the world of sense and at the same time to the intelligible world, are no longer thought merely indeterminately and problematically (which even speculative reason could detect as possible), but with respect to the law of its causality are determinately and assertorically known (*erkannt*); thus is the reality of the intelligible world definitely established from a practical point of view, and this determinateness, which would be transcendent (extravagant) for theoretical purposes, is for practical

I think that the general outline of Kant's treatment of this problem is defensible.

Arguably, moral responsibility requires the sort of freedom he believes it does, while our being free in this sense cannot be established on the evidence. Nevertheless, there is a consistent conception according to which we have freedom of this kind. Questions arise about the credibility that Kant requires for this conception, and about the practical reasons he adduces for the belief that we are free in the sense at issue. I argue that the credibility issue resolves in Kant's favor, while the practical reasons he cites are subject to a serious challenge. But this last issue is complicated, and an overall view of the sort Kant develops remains significant and, in key respects, attractive.

1. The outline of the problem Kant sets out for free will 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

---

purposes immanent. (*KpV*, *Ak* V 105, from Lewis White Beck's translation of Kant's *Critique of Practical Reason* (Indianapolis: Bobbs-Merrill, 1956.)

Ameriks observes about these sorts of claims in the *Critique of Practical Reason* that "however true it may be that the moral law gives us a way to think our freedom more concretely, strictly speaking the idea of the law as such can add only to the content of that thought and not its theoretical certainty," p. 219). Robert Adams suggests (with an acknowledgment to Jessica Moss) that Kant holds that our experience of moral obligation provides content to our idea of transcendental freedom by giving us a sense of what it would be like to be transcendently free ("Things in Themselves," *Philosophy and Phenomenological Research* 57 (1997), pp. 801-25, at p. 19). Perhaps this is because our experience of being obligated to, say, refrain from something that we are doing gives us a sense of being able to act otherwise than how we actually do.

would be no human action that we could not predict with certainty, and recognize as necessary given its preceding conditions. (A549-50/B577-8)

At the same time, Kant believes that this fact does not rule out the claim that some of our actions are free. Yet especially in the *Critique of Practical Reason*, he emphatically rejects any compatibilist solution according to which free action is compatible with its causal genesis being *exhausted* by preceding natural conditions that causally determine its occurrence. There he contends against Hume that compatibilist freedom of this sort is ruled out for the following reason:

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)<sup>4</sup>

---

<sup>4</sup> It is sometimes held that Kant's only reason for thinking that phenomenal determinism poses a problem for the sort of freedom required for morality and moral responsibility is that he believed that phenomenal determinism entails determinism by sensuous impulses. But this passage indicates that he has another concern – that phenomenal determinism entails determinism by the past, and determinism by the past undermines the sort of control required for the sort of freedom at issue. This passage also presents a challenge to the interpretation of Hud Hudson, who follows Ralf Meerbote in reading Kant as a compatibilist of a Davidsonian sort (*Kant's Compatibilism*, esp. pp. 39-56; cf. Ralf Meerbote, "Kant on the Nondeterminate Character of Human Actions," and Donald Davidson, "Mental Events," in *Experience and Theory*, L. Foster and J. Swanson, eds. (London: Duckworth, 1970), reprinted in Donald Davidson, *Essays on Actions and Events* (Oxford: Oxford University Press, 1980), pp. 207-25). In Hudson's view, an event in the empirical world, which is correctly described as causally determined, can also be a free action by virtue of being token-identical to an event that can correctly be described as resulting from a pro-attitude, propositional determination -- a description that does not involve the concept of causal determination or related notions. By Hudson's characterization:

In the case of what we might call pro-attitude, propositional determination (what Kant would term "the determination corresponding to an intelligible cause") if an agent S in performing action x is determined by something y, then y is S's practical reason for performing x, and y consists in the conjunction of a desire for some end and a belief in a proposition expressing means to that end. Whereas such determination is necessary but not sufficient for its corresponding action ... that is, whereas some agent's having a

Moreover,

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*)

For Kant, it is a requirement of free action that its causal genesis not be exhausted by preceding natural conditions that causally determine its occurrence, for only then could it be that the action is in the subject's power in a sense sufficient for practical freedom.

[Practical freedom] "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

---

practical reason for performing some action is not invariably followed by a performance of that action, the sense of 'determination' at work in the Second Analogy differs precisely in this regard. (pp. 42-3)

Hudson interprets Kant's notion of the causality of reason as this sort of determination, which is not causal determination. However, if this is all that causality of reason amounts to, then Kant's theory fails to respond to the concern expressed in the passage just quoted -- that on theories such as Hume's, because the past is not in my control, every action that I perform must be necessary by determining grounds that are not within my control. Terence Irwin has urged, and Hudson admits, that if an event is determined by preceding conditions, it is true that it is determined by preceding conditions under all descriptions (Terence Irwin, "Morality and Personality: Kant and Green," in *Self and Nature in Kant's Philosophy*, ed. Allen Wood (Ithaca: Cornell University Press, 1984), pp. 31-56, at p. 38; see Hudson's discussion on pp. 82ff.). I believe that Irwin's claim can be challenged with some success -- see my discussion of the dream analogy in what follows. Still, escaping Irwin's concern requires more idealism than Hudson countenances. Hudson suggests that Kant's main worry about determinism is that our actions would be determined by non-rational motivating factors, such as sensuous impulses. But again, as this passage indicates, Kant is also concerned about determination by the past, for the reason that in his view such determination undermines the kind of control at issue.

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)<sup>5</sup>

Practical freedom, in its negative aspect, is the will's power to act (or to choose to act) without being causally determined by sensuous impulses (A534/B562), and in its positive aspect it is the will's power to act motivated by principles whose source is not in sensuous impulses but rather in rationality itself (e.g., *KpV*, *Ak V*, 129). Here Kant indicates that in his view practical freedom presupposes a power to act independently of the natural causes that determine our actions.<sup>6</sup>

What kind of indeterministic causation of actions is required for them to be free? Like Hume, Kant rejects the notion that free actions could simply be indeterministically caused events in nature, as in the view of Lucretius – such events would amount to “blind chance” (*KpV*, *Ak V* 95).<sup>7</sup> Rather, the sort of causality required is “the power of beginning a state *of itself* (*von selbst*)

---

<sup>5</sup> This passage specifies a notion of independence of natural causes that freedom does not have on Hudson's reading (see the previous note). Here is Hudson's interpretation of independence: “Determination, so considered, is our pro-attitude, propositional determination, and Kant's claim that this is independent of causal determination can now be read as follows: an imperative can determine the will in the sense of providing the propositional component of a practical reason for the agent's action, and such pro-attitude, propositional determination is not expressed with causal, determining descriptions” (*Kant's Compatibilism*, p. 48). However, this kind of independence is compatible with the absence of a power to produce actions independently of natural causes and even opposed to their power and influence. Moreover, I don't see how a compatibilism of the sort Hudson develops could accommodate the sort of independence that Kant specifies here.

<sup>6</sup> Wood appropriately emphasizes that for Kant freedom is a *power*; “Kant's Compatibilism,” pp. 79-83.

<sup>7</sup> The passage reads: “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

– 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). Kant classifies this power as an “*intelligible cause*” (e.g. A537/B565).<sup>8</sup> In this conception, a free action is not merely an uncaused event that occurs in the agent. Rather, when an action is free, the agent *of itself* produces an act in the sensible world, and, in producing this act of itself, the agent is not determined to do so by preceding causes (A541/B569). Kant calls this characteristic of agents *transcendental freedom*.<sup>9</sup>

One element in Kant’s definition of our *power of free choice* (*Willkür*) – not the same notion as transcendental freedom -- is “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). One

---

set out in his *De Rerum Natura*, tr. W. H. D. Rouse, Loeb Classical Library (Cambridge: Harvard University Press, 1982), 2.216-293: “...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” (2.289-293);

<sup>8</sup> It has often been charged that Kant’s invoking noumenal causality here violates his own epistemic strictures of the knowledge of causation that he argued for earlier in the *Critique of Pure Reason* (see for example, Jonathan Bennett, *Kant’s Dialectic* (Cambridge: Cambridge University Press, 1974), pp. 189-95). But Robert Adams points out that Kant allows that we be able to *think* noumena (B166, A254/B309), and noumenal causation in particular, and that the thought of noumenal causation has a central role in Kant’s practical philosophy (“Things in Themselves,” p. 820). Adams suggests, in addition, that “it is plausible, from a Kantian point of view and for theoretical purposes, to regard the concept of noumenal causality as a problematic concept. It is a concept of a real (and not merely logical) relation that corresponds to the inferential form (and force) of the hypothetical judgment, and that is not understood in terms of succession of events in time, nor in any other terms that depend on our forms of intuition” (p. 821).

<sup>9</sup> In the contemporary debate, the advocates of transcendental freedom are typically agent-causal libertarians, who hold that the agent fundamentally as substance is the transcendently free cause. Eric Watkins, in *Kant and the Metaphysics of Causality* (Oxford: Oxford University Press, 2005, pp. 230-97) argues that for Kant causation is always most fundamentally substance-causation. If this is right, causation by the power of transcendental freedom would in this respect be like any other causation.

might presume that transcendental freedom also essentially involves the ability to do otherwise.

But a passage in the *Religion Within the Boundaries of Mere Reason* intimates that transcendental freedom need not involve this ability:

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 morally good or right. Still, God is free – and presumably, transcendently free – by virtue of the fact that God is absolutely spontaneous in the production of action. That is, the determining ground for action lies solely within the divine self, which entails that action is not determined by preceding conditions. This suggests that transcendental freedom does not essentially involve the ability to do otherwise. Kant would seem to be a *source* rather than a *leeway* incompatibilist, stressing that the key notion of freedom is not the ability to do otherwise, but rather being the undetermined source of one's actions.<sup>10</sup> At the same time, as

---

<sup>10</sup> I first argued for source incompatibilism in "Determinism *Al Dente*," *No φs* 29, 1995, pp. 21-45, and subsequently in *Living Without Free Will* (Cambridge: Cambridge University Press, 2001), chapters 1 - 4. Others who have argued for an incompatibilism of this kind include Robert Heinaman, "Incompatibilism without the Principle of Alternative Possibilities," *Australasian Journal of Philosophy* 64 (1986), pp. 266-76; Eleanore Stump, "Intellect, Will, and the Principle of Alternate Possibilities", in *Christian Theism and the Problems of Philosophy*, ed. Michael Beaty, University of Notre Dame Press, 1990, pp. 254-285; Linda Zagzebski, *The Dilemma of Freedom and Foreknowledge* (New York: Oxford University Press, 1991), Chapter

we shall see in Section 5, it is of great significance for Kant that we human beings have the ability to do otherwise, since this is a necessary condition for ‘ought’ principles applying to us, and for us the moral law is a system of ‘ought’ principles. But for Kant there is no corresponding reason to claim that God can do otherwise, since God is not subject to ‘ought’ principles; ‘for the divine will... there are no imperatives: ‘*I ought*’ is here out of place, because ‘*I will*’ is already of itself in harmony with the law” (*G*, Ak IV, 414). In fact, in Kant’s view God cannot do otherwise, but God is still free in the sense he has in mind.

Allen Wood takes Kant to hold that an agent’s will can be transcendently free as long as her actions are not determined by temporal causes – that is, by preceding natural conditions -- and, furthermore, its being transcendently free is compatible with the necessitation of her actions by her nature. In support, Wood contends: “Kant holds that a holy will is free even though its acts are necessitated, because they are necessitated from within by reason rather than by the sensuous impulses that are foreign to our nature.”<sup>11</sup> However, Kant denies that God, whose will is holy, is necessitated by his nature when he acts:

One might raise the objection that God cannot decide otherwise than he does, and so does not act freely but out of the necessity of his nature... But in God it is not due to the necessity of his nature that he can decide only as he does. Rather it is true freedom in

---

6, Section 2.1; “Does Libertarian Freedom Require Alternate Possibilities?” *Philosophical Perspectives* 14 (2000); Michael Della Rocca, “Frankfurt, Fischer, and Flickers,” *No  s* 32 (1998), pp. 99-105; David Hunt, “Moral Responsibility and Unavoidable Action,” *Philosophical Studies* 97 (2000), pp. 195-227. All of these incompatibilists are motivated to source incompatibilism by their acceptance of arguments from Frankfurt-style examples against a principle of alternative possibilities.

<sup>11</sup> “Kant’s Compatibilism,” p. 82.

God that he decides only what it suitable to his highest understanding. (*VpR, Ak XXVIII*, 132, Wood translation, pp. 105-6).

On my reading, transcendental freedom is incompatible with being necessitated or determined to act by one's nature.

The sense of Kant's overall position on transcendental freedom is this. If the causal genesis of an action is exhausted by preceding natural conditions, or if an action is simply an indeterministically caused event, then the agent will not have the control over it that genuinely free action demands. On the deterministic conception, the requisite sort of control is lacking because factors beyond the agent's control are sufficient to produce the action. Kant does not say exactly why Lucretius's indeterministic alternative is incompatible with free action, but in this conception, antecedent conditions leave it open whether the action in question will occur, and given the causal contribution of these antecedent conditions, the agent has no further role in determining which of the options is taken. It would be in the spirit of Kant's view to claim that the agent's having no further role at this point entails that she has insufficient control for free action. The further role the agent must have is causal. Transcendentally free agents have the causal power to produce an action, and thus to determine that an action takes place, without being causally determined by antecedent conditions, nor by their natures, in their exercise of this power. The proposal, then, is that the control absent in both the deterministic and indeterministic conceptions is supplied by the agent's power of transcendental freedom.<sup>12</sup> Having the power of

---

<sup>12</sup> When an agent A causes decision D at t by her transcendental freedom, then an event of the following type occurs: A's causing D at t – call this event 'E.' One might object that given exactly the same antecedent conditions as those that precede A's agent-causing D, E might not have occurred, so that E occurred would still seem to be just as much a matter of chance as a free action would on Lucretius's view; (Carl Ginet raises such an objection in "Freedom, Responsibility, and Agency," *The Journal of Ethics* 1 (1997), pp. 85-98, at p. 91). The advocate

transcendental freedom does not essentially involve the ability to do otherwise, but human beings nevertheless often have this ability.

2. But now, given Kant's empirical determinism about human action, and this apparently indeterministic conception of free action, how can he avoid hard determinism – the view that because all of our actions are causally determined they are not free? Kant contends that a reconciliation between empirical determinism and transcendental freedom might be provided by transcendental idealism.<sup>13</sup> First, acting human subjects, as appearances, have an *empirical character*. Character, Kant says, is a *law of something's causality*, where causality is the activity

---

of transcendental freedom must indeed admit that given the relevant antecedent conditions, E might have occurred or not. But the key issue is whether the agent nevertheless can have the crucial role in determining whether or not the decision occurs that she cannot have on Lucretius's view. And it seems to me that she can. What the transcendently free agent does most fundamentally is to cause an action, or perhaps more precisely, a decision, *from herself*, qua agent. At this point, one should note that it is a logical consequence of the transcendently free agent's causing a decision that an event of E's type occurs. It follows logically from the fact that Ann causes the decision to flip the coin that the event Ann's causing the decision to flip the coin at t occurs. But it is by transcendently freely causing a decision that the agent brings about the event of type E - as a logical consequence of her causing the decision. What thus explains the occurrence of the event of type E - indeed, already given the causal contribution of the antecedent conditions - is Ann, qua transcendently free agent, causing the decision from herself. This contrasts with Lucretius's scenario, where given the causal contribution of the antecedent conditions, the agent has no further role in determining whether the decision occurs, and so events of type E, will, given the causal contribution of the antecedent conditions, have no further explanation, and, in particular, no further explanation involving the agent. Agent causation, more generally, can solve the problem for free action that Lucretius's position faces. (See my *Living Without Free Will*, pp. 38-68.)

<sup>13</sup> This part of his position has been nicely laid out by Allen Wood, "Kant's Compatibilism," pp. 83-9, and Henry Allison, *Kant's Theory of Freedom*, pp. 30-41. A feature of Allison's account that I find particularly convincing is his argument that empirical character is best viewed as an appearance or manifestation, and not simply a result, of intelligible character (p. 32).

of a cause (A539/B567). We might say that a thing's character is the way it behaves causally. A key feature of how we as appearances or empirical subjects behave causally is that when we act we are determined by preceding natural conditions. But, Kant argues, this empirical character is compatible with our actions in the empirical world being produced by virtue of a character of a very different sort. It may be, in particular, that as a thing in itself or noumenon the agent has an *intelligible* character (A539/B567), whereby in its production of an action it is transcendently free and thus not causally determined by preceding natural conditions.

Kant does not claim that we can *know* that we as intelligible or noumenal subjects are transcendently free; "we have not been trying to establish the reality of freedom." In fact, he claims that "we have not even tried to prove the possibility of freedom." Rather, he hopes only to establish that "nature at least *does not conflict with* causality through freedom" (A558/B586). In his terminology, he does not aim to show that our concept of transcendental freedom is *really possible*, but only that this concept is *logically possible*. In Kant's view, for a concept to be logically possible is for that concept itself not to feature a contradiction (A244/B302). How does real possibility differ? The 'real' in 'real possibility' indicates, as Robert Adams points out, the notion of '*reales*,' or realities, which are positive (in the sense of non-negative) features (e.g., A602/B630).<sup>14</sup> At the end of the discussion of the ontological argument in the *Critique of Pure Reason* Kant claims that although the concept we have of God gives rise to no contradiction, nonetheless we cannot ascertain the possibility of the existence of God, whereby he means its real possibility. There he cites two ways in which such knowledge might be precluded: either by the realities not being "given to us specifically," or by our inability to determine whether the

---

<sup>14</sup> Robert Adams, "Things in Themselves," pp. 813-6.

realities can be connected in a thing (A602/B630). Given these clues, it seems reasonable to conclude that a concept is really possible when it is *metaphysically possible for the thing of which it is a concept* to exist, and that this is at least in part a matter of whether it is metaphysically possible for the positive properties that the thing would feature to exist, and of whether it is metaphysically possible for those positive properties to combine into the thing in question.

Kant argues that *we* can know (*erkennen*) the real possibility of a concept theoretically only through experience, and in particular by sensible intuition (A602/B630; cf. A218/B265-A226/B274). Given the sort of cognitive equipment we have, we can only determine by means of experience whether the conditions of real possibility are satisfied. This fact limits us to discerning real possibilities only for objects of experience, and not for noumena. So, although our concept of transcendental freedom contains no contradiction, this is not enough to show that this sort of freedom satisfies the conditions of real possibility. On the hypothesis that transcendental freedom would be a simple causal power, and thus not constituted by a complex of underlying causal powers, it might be that a simple causal power of this kind is not metaphysically possible. Or, supposing that transcendental freedom would be constituted of a complex of underlying causal powers, some of the positive properties that it would have to feature might nevertheless not be metaphysically possible. Or otherwise, supposing that these constituting causal powers are all metaphysically possible, it might not be metaphysically possible for these positive properties to combine into the power in question. At the same time, we will never be able to determine whether any of this is in fact so.

Elsewhere I have contended that when Kant denies knowledge of noumena or things in

themselves, he is concerned in particular to reject what I call *substance-knowledge* of them.<sup>15</sup> Substances in the Leibnizian scheme – which Kant takes over in his conception of things in themselves – are entities whose essential features are fundamental causal powers, and these powers are conceived as intrinsic properties of these entities. For Leibniz, all of the substances are monads, whose essential feature is the fundamental causal power of representation, and this power of representation the monad has intrinsically. In denying knowledge of things in themselves, Kant is ruling out our grasp of such fundamental causal powers of substances. This account of the denial of knowledge of things in themselves, I have argued, solves the problem that many commentators, including P. F. Strawson, raise for Kant’s transcendental philosophy: that he rules out knowledge of things in themselves, but at the same time his transcendental philosophy’s claims about the unity of apperception, synthesis, and the categories are surely claims about things in themselves, rather than appearances, and hence Kant’s entire enterprise is deeply inconsistent.<sup>16</sup> For if the knowledge of things in themselves that is precluded is just our grasp of these fundamental noumenal causal powers, then the inconsistency may well vanish. For arguably, transcendental philosophy makes no knowledge-claims about the nature of fundamental noumenal causal powers – of the self as it is in itself, in particular. Indeed, after providing his exposition of transcendental philosophy in the *Critique of Pure Reason*, Kant affirms that nevertheless “the relation of sensibility to an object, and what might be the

---

<sup>15</sup> "Is Kant's Transcendental Philosophy Inconsistent?" *History of Philosophy Quarterly* 8, 1991, pp. 357-372; cf. Rae Langton, *Kantian Humility* (Oxford: Oxford University Press, 1998).

<sup>16</sup> P. F. Strawson, *The Bounds of Sense* (London: Methuen, 1966).

transcendental ground of this unity, undoubtedly lie too deeply hidden for us, who know even ourselves only through inner sense, thus as appearance, to be able to use such an unsuitable tool of investigation to find out anything except always more appearances, even though we would gladly investigate their non-sensible cause” (A278/B334).

Transcendental freedom would indeed be a fundamental causal power – an intrinsic feature of human selves as they are in themselves, and so knowledge of this power is ruled out by Kant’s stricture. Thus, in accord with this account, although we can form a superficial conception of transcendental freedom by means of our reason, 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. Transcendental freedom could in fact turn out to be metaphysically impossible, or the nature of fundamental noumenal causality might actually not allow for transcendental freedom, but this we could never discover.

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 really impossible, or the fundamental nature of noumenal causality precludes transcendental freedom? The best interpretation of Kant’s claim here is that there is no internal inconsistency in the superficial description 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 understand this power, that is, we could never comprehend its nature as a fundamental power, and that we could never understand the nature and arrangement of whatever complex structure of fundamental causal powers would underlie it.

Kant's position is that transcendental freedom is in one sense *conceivable* and in another sense not. It is instructive to locate this position among several of the dimensions of conceivability outlined by David Chalmers. By Chalmers's characterization, S is *prima facie* conceivable when S is conceivable on first appearances, and *ideally* conceivable when it is conceivable on ideal rational reflection.<sup>17</sup> Second, S is *negatively* conceivable in general when S is not ruled out, and S is negatively conceivable in the central sort of way when S is not ruled out a priori. *Positive* notions of conceivability require that one can form some sort of positive conception of a situation in which S is the case.<sup>18</sup> In Kant's view, transcendental freedom is negatively conceivable given our cognitive situation – we cannot rule it out. This is consistent with its not being ideally negatively conceivable. I think that Kant should have said that due to the possible limitations of our powers of reasoning, we are in no position to know whether transcendental freedom is or is not ideally negatively conceivable. Kant does not make these concessions, but I think he should have. But he would agree that transcendental freedom is in a sense not now positively conceivable by us, since we cannot form a positive conception of the fundamental causal powers that would constitute it. However, it may or may not be ideally positively conceivable – we are in no position to know which. Accordingly, we are in no position to know whether transcendental freedom is metaphysically possible, and this is a conclusion that Kant does affirm.

---

<sup>17</sup> David Chalmers, "Does Conceivability Entail Possibility?" in *Conceivability and Possibility*, Tamar Szabo Gendler and John Hawthorne, eds (Oxford: Oxford University Press, 2002), pp. 145-200, at p. 147.

<sup>18</sup> David Chalmers, "Does Conceivability Entail Possibility?" pp. 149-50.

3. At this point it is important to distinguish among several aspects of Kant's philosophical aims for transcendental freedom.<sup>19</sup> Two are prescriptive, a third is descriptive. First, there is the investigation regarding the epistemic or theoretical rationality of the belief that we are transcendently free. As a matter of epistemic rationality, the Kantian position is that we can vindicate only the internal and external consistency of this belief. Accordingly, Kant would not aim to convict of epistemic irrationality the skeptic who denies that we can epistemically justify the belief that we are transcendently free. Second, Kant strives to show that it is practically rational for us to believe that we are transcendently free, and, indeed, practically irrational for us not to. He does intend to show that the skeptic about transcendental freedom is practically irrational. Third, in addition to these two prescriptive aims, there is also the descriptive Kantian project of developing a moral theory in which rationality, autonomy, and transcendental freedom all have a part, a theory that would successfully rival competitors such as Hume's. What follows does not challenge this descriptive project so much as the second prescriptive aim. I shall argue that it is not clear that the practical reasons Kant adduces provide practical justification for the belief that we are transcendently free. At the same time, what I have to say on this score is relevant to the descriptive project, for I believe that Kantian ethics, conceived more broadly, can flourish without the belief in transcendental freedom. But I will not pursue that issue here.

In Kant's view, we cannot know the fundamental causal powers of things in themselves, but we can nonetheless have legitimate "beliefs" about these causal powers. The relevant notion of belief in this context is a "subjectively sufficient" but "objectively insufficient" conviction that is based on practical and not on theoretical – i.e. evidential – considerations (A822-3/B850-

---

<sup>19</sup> The points in this paragraph were prompted by a referee for *Philosophy and Phenomenological Research*.

1).<sup>20</sup> Kant also says that to believe in the practical sense is *just* to be guided practically by the content of the belief; the word ‘belief’ in this context “concerns only 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 one sets aside any reservations one might have about the truth of the proposition to be believed – reservations that may result from inadequate evidence -- so that one’s conviction in that proposition (or another proposition relevantly related to it<sup>21</sup>) can guide one’s thought and deliberation about acting.

Kant maintains that for some such beliefs, although their theoretical basis is inadequate, the practical benefits that derive from their capacity to guide action justify our having them. Thus, for example, he argues that for these sorts of practical reasons one is justified in holding the belief that God exists despite its lacking adequate theoretical grounding. At the same time, he indicates that in general such beliefs – at least when they are about things in themselves – should still meet a theoretical requirement: they must not be internally inconsistent or inconsistent with anything we know. Why does he advocate this requirement? On the one hand, it is clear that Kant maintains that the law of non-contradiction governs things in themselves.

---

<sup>20</sup> It should be emphasized that the notion of belief that Kant has in mind here involves conviction, and belief is to be distinguished from opinion: “If it were possible to settle by any sort of experience whether there are inhabitants of at least some of the planets that we see, I might well bet everything that I have on it. Hence I say that it is not merely an opinion but a strong belief (on the correctness of which I would wager many advantages in life) that there are also inhabitants of other worlds. (A825/B853).

<sup>21</sup> As we shall see, what may sometimes be at issue is belief in a proposition that does not itself provide practical guidance, but nonetheless would support a proposition that does.

But still, if the practical rationality of belief is at issue, then it would seem that not even inconsistency should in principle rule it out. Suppose someone with a cerebroscope seriously and credibly threatened to torture and kill me unless I believed the proposition *that some other planet is inhabited by intelligent life and no other planet is inhabited by intelligent life*, or, say, Russell's paradox. Might it not then be practically rational to believe such a proposition (for instance, if I could induce this belief temporarily by taking a drug that has no other bad consequences)? Or suppose that in order to maintain that moral principles hold for us we would have to believe some inconsistent proposition. Shouldn't Kant at least entertain the possibility that believing the inconsistency is legitimate on practical grounds?

Perhaps he would entertain this possibility – Kant doesn't explicitly rule on the issue. But his concern is rather that one can be certain that an inconsistent proposition is false, and if one is certain that a proposition is false, then it won't be psychologically possible to believe it. In the Canon of Pure Reason, 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 inconsistency, then their impossibility would be provable. Then we could be certain that these claims were false, and if we

were, we could not have the conviction in them required to secure the relevant practical effect. To avoid the certainty of the falsehood of these claims for those who would become aware of an inconsistency, it would have to be shown that they did not have this feature. Kant takes a bolder stance in a passage in the *Critique of the Power of Judgment*, where he argues that in order to maintain the possibility of the highest good (happiness in accord with virtue), or to avoid the weakening of one's respect for the moral law that would result from not believing that the highest good will be realized, an agent must believe that there is a God: "he must assume the existence of a moral author of the world, i.e. of God, from a practical point of view, i.e. in order to form a concept of at least the possibility of the final end that is prescribed to him by morality – *which he can very well do, since it is at least not self-contradictory*" (KU, Ak V, 452-3, emphasis mine). Here Kant seems to be asserting that the absence of a contradiction in a proposition actually makes it possible to assume that it is true.

However, it is also clear that propositions can lack credibility in the sense at issue for reasons other than overt inconsistency. A proposition's being obviously very highly improbable might render it at most insignificantly more credible in this sense than an overtly inconsistent proposition. For example, consider the proposition that Martians will attempt to take over the earth tomorrow, which for me is highly improbable. This proposition features no inconsistencies, but I nevertheless cannot set aside my reservations about its truth so that my conviction in it can guide my thought and deliberation about acting. At the same time, for the purposes of this discussion, I don't want to set the standard of credibility too high – as, for example, credibility for any rational person, regardless of philosophical persuasion. If Kant's conception of freedom requires only philosophical views that are not uncommonly accepted by people who have

considered them ably and seriously, it will count as sufficiently credible.

4. So even if the belief that we are transcendently free does not involve an overt contradiction, is it significantly more credible in this sense than a proposition that does? A first issue that bears on the issue of credibility is whether the empirical subject who is causally determined to cause the action is to be understood as *identical* to the noumenal subject who 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*)

The majority of interpreters favor a *one-world* view on the relationship between phenomena and noumena, according to which in general an appearance-object is identical to a noumenon. It is often noted that there are several texts that support the one-world view, while there are also those that might be construed to favor the *two-world* reading, according to which the appearance and the noumenon are distinct.<sup>22</sup> Now it might well seem difficult to see how the belief that we are transcendently free would avoid incredibility on a one-world conception. Imagine an action in the empirical world that an empirical subject causes. An empirical subject is, in Kant's view, a complex of psychological states that can be apprehended by inner sense. By his account, this empirical subject is causally determined to perform the action by preceding conditions. Consider

---

<sup>22</sup> E.g. James van Cleve, *Problems from Kant*, (New York: Oxford University Press, 1999), pp.

the hypothesis that there is a noumenal subject that causes the action from itself without being causally determined by preceding conditions. Could the empirical subject E and this noumenal subject N be identical? The empirical and the noumenal subjects differ in their properties: E has the property of being determined by preceding conditions to cause the action, while N does not have this property. And if E has a property that N lacks, then E and N are not identical – at least one would think this was so pending further explanation. More specifically, 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?<sup>23</sup>

A second issue that bears on the question of credibility is that at least on the two-world reading, it would seem that the action, an event in the empirical or phenomenal world, is overdetermined in a peculiar way. By one strand of its causal history this empirical action has a sufficient cause in a transcendently free subject, while by another strand this same action has a sufficient cause in a deterministic series of events that traces back to a time before the (empirical) agent was born.<sup>24</sup> To see the problem, consider first the suggestion that these two sorts of causal

---

<sup>23</sup> A view has recently emerged according to which there is nothing problematic in maintaining that our actions are both free and determined for the reason that these beliefs are held from different standpoints, one practical and the other theoretical, each bringing with it its own standards for rationality. See, for example, Christine Korsgaard, “Morality as Freedom, in *Kant’s Practical Philosophy Reconsidered*, ed. Y. Yovel (Dordrecht: Kluwer, 1989). For a critical discussion of the claim that the practical and theoretical standpoints can be integrated in this way, see Dana Nelkin, “Two Standpoints and the Belief in Freedom,” *Journal of Philosophy* 97 (2000), pp. 564-76.

<sup>24</sup> A reviewer objects as follows: Here I illegitimately assume that for Kant the world of appearances can contain the sufficient conditions for a given event. But Kant maintains that this is not so; he explicitly argues in the Third Antinomy that there is no completeness in the series of causes, from which he infers that something would “happen without a cause sufficiently determined a priori” (A446/B474).

strands might be reconciled without the distinction between phenomena and noumena that Kant advocates. Now there would certainly 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, Kant needs a much more substantial proposal. 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. It might initially appear that the wild coincidences implied by this proposal make it quite incredible. If we were agents making transcendently free choices for our actions, would we not expect, in the long run, that these choices be evident in the world as patterns of divergence from the deterministic natural

---

In response, at A445-6/B473-4 Kant is clearly claiming that the “sufficient determination a priori” of event E at T0 requires not just that:

(1) There exist conditions C at some time T1 prior to T0 that are causally sufficient for E but something stronger, reflected in the following passage:

“The causality of the cause through which something happens is always something that has happened, which according to the law of nature presupposes once again a previous state and its causality, and in the same way a still earlier state, and so on.” (A445/B473)

So for event E to be “sufficiently determined a priori” it must be that:

(2) There exist conditions C1 at some time T1 prior to E that are causally sufficient for E, and there exist conditions C2 at some time T2 prior to T1 that are causally sufficient for C1, and there exist conditions C3 at some time T3 prior to T2 that are causally sufficient for C2, etc..

Now for any empirical action, Kant believes that (1) holds, as the following passage indicates:

... 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/B577-8)

Given our use of the term ‘causally sufficient,’ we can conclude from this passage that Kant holds that every empirical action has preceding empirical conditions that are causally sufficient for it -- even if this action is not “sufficiently determined a priori.” Supposing that an action also has a sufficient noumenal cause, we can also conclude that in a familiar sense the action will be overdetermined, by virtue of a sufficient noumenal and a sufficient phenomenal cause.

laws? Wouldn't the proposal that there are no such divergences, despite involving no contradiction, run so sharply counter to what we would expect to occur as to render the proposal incredible – at best insignificantly more credible than an outright contradiction?<sup>25</sup> But perhaps a further story can be told to make this proposal credible, and this is indeed what Kant aims to do.

Let us first consider the question of incompatible properties. On Allison's one-world theory, the empirical subject E, *considered under the conditions of the possibility of experience*, is causally determined to cause the action by preceding conditions, while that same subject, *considered independently of these conditions*, now the noumenal subject N, produces the action of itself while not being causally determined in this way.<sup>26</sup> What Allison's theory now requires is an explanation as to how it could be that I am causally determined to perform an action when I am considered under the conditions of the possibility of experience, while I am not causally determined when I am considered independently of those conditions.<sup>27</sup> Or focusing on the action rather than the agent, it could be that an action, considered empirically, is causally determined,

---

<sup>25</sup> One might think, in accord with some recent agent-causal libertarians, that this problem would dissolve if Kant had concluded in the Second Analogy that it would be sufficient if the empirical world were governed by statistical and thus indeterministic rather than deterministic laws. I argue that this proposal fares no better – *Living Without Free Will*, pp. 79-85

<sup>26</sup> Henry Allison, "Transcendental Idealism: The Two Aspect View," in *New Essays on Kant*, Bernard den Ouden and Marcia Moen, eds. (New York: Peter Lang, 1987), pp.158-78; *Kant's Theory of Freedom*, pp. 3-5.

<sup>27</sup> Allison attempts to explain how transcendental idealism solves the incompatible properties problem, *Kant's Theory of Freedom*, pp. 41-6. One might object that indexing the properties to different perspectives, or else, to different realms, all by itself solves the incompatible properties problem. But this is not so, for the reason that the different perspectives or realms are nevertheless related in certain ways (as illustrated by the present discussion), and the incompatibility of the properties needs to be consistent with these relations.

while that same action, considered noumenally, is not. Here Terence Irwin's challenge is especially pertinent: "if an event is determined, it is true of it under all descriptions that it is determined, even though only some true descriptions, those referring to the relevant laws, show why it is determined."<sup>28</sup> What resources does Allison's view have to answer Irwin?

Allison's general one-world but two-aspect attempt at reconciliation has been influential, but some commentators have found this theory puzzling. James van Cleve, for example, seeks but finds no ordinary analogy that would show how it can resolve the problems for incompatible properties Kant's view faces.<sup>29</sup> But here we might make use of an analogy with dreams that Kant himself suggests in his "transcendental hypothesis" designed to defend the belief that we are immortal:

...that all life is really only intelligible, not subject to temporal alterations at all, and has neither begun at birth nor will be ended through death; that this life is nothing but a mere appearance, i.e., a sensible representation of the purely spiritual life, and the entire world of the senses is a mere image, which hovers before our present kind of cognition and, like a dream, has no objective reality in itself; that if we could intuit the things and ourselves *as they are*, we would see ourselves in a world of spiritual natures with which our only true community had not begun with birth nor would cease with bodily death (as mere appearances)... (A779-80/B807-8)

Suppose I dream that you fly to Aruba. Then it might be said that you, considered in the context of the dream, have the property of flying to Aruba, while you, considered independently of the

---

<sup>28</sup> Terence Irwin, "Morality and Personality: Kant and Green," p. 38.

<sup>29</sup> James van Cleve, *Problems From Kant*, pp. 146-50.

dream context, lack this property. We could invent an *in-dream-D* operator to capture this idea. You really lack the property of flying to Aruba, but under the scope of the in-dream-D operator you have the property of flying to Aruba. Alternatively, what you have is a property that is a relation to my dream: perhaps the property of *being dreamt by me to be flying to Aruba*. In the dream case, apparently incompatible properties can be rendered compatible by either of these mechanisms.

So if the relation between appearances and things in themselves is relevantly similar to the relation between dreams and waking life, then we have a strategy for defusing the problem of incompatible properties for the one-world view. I lack the property of being causally determined by preceding conditions to produce my action, but I, under the scope of an as-appearance operator, have this property; or alternatively, I have the property of appearing as causally determined.<sup>30</sup>

---

<sup>30</sup> Kant seems to claim much more generally that empirical objects have properties that things in themselves lack, and vice-versa. For example, empirical objects have spatial qualities such as squareness, whereas things in themselves lack such qualities. More generally, all empirical properties could be construed relationally: to say that the earth is spherical, more explicitly, is to say that it has the property of *appearing as spherical*. This last idea is familiar, but it is the dream analogy that allows us to see how one can profitably understand it as a key feature of the one-world position. Now I think that van Cleve is right to say that it would be foreign to Kant to assert that spatial properties are relational in this way. It is a feature of Kant's empirical realism that objects of experience straightforwardly have spatial properties. However, I suspect that in the case of physical objects this is the best we can do for the one-world sentiments he frequently professes. But it may also be, as Adams suggests, that Kant does not think of individual physical objects as identical to or as even as corresponding to individual things in themselves at all, while in the case of human subjects he does think of the identity or correspondence as holding (in "Things in Themselves"). Especially in the Amphiboly, Kant models things in themselves on Leibniz's system of monads, and for Leibniz there is no one-to-one correspondence between individual macroscopic physical objects and individual monads. At the same time, we have already encountered passages in the *Critique of Practical Reason – Ak V*, 9 and *Ak V*, 97 -- in which Kant asserts the identity between the empirical and the noumenal subject. Human subjects, as Kant conceives them, do not have spatial properties, so here this particular worry of van Cleve's does not arise.

Perhaps this dream analogy might be filled out to address our more specific concern about the plausibility of Kant's way of reconciling freedom and empirical determinism given the one-world assumption: (the problem being that) 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 particular way? Here an embellished version of the dream analogy is instructive. Suppose I am capable of so-called lucid dreaming, of controlling by my will the contents of my dreams, and imagine that this control is the transcendently free sort. One night I freely will myself to dream that I am deterministically manipulated by sophisticated Martian neuroscientists to steal the Mona Lisa. In these circumstances, I freely cause the dream-action of stealing the Mona Lisa, while at the same time, I, in this dream, am causally determined to carry out the dream-action by the neuroscientists. Furthermore, in this example, by way of reply to Irwin, it turns out that an event – say, the occurrence of the dream-content *my stealing the Mona Lisa* – is causally determined considered or described as a dream occurrence, but is not causally determined (by anything other than the transcendently free agent) considered or described as an event in the real world.

On Kant's conception, are my appearance-producing capacities sufficiently similar to my dream-producing capacities for this analogy to help resolve the credibility issue for his account of freedom? A specific worry is that, by contrast with the lucid dreaming case, I, as a thing in itself, do not have a sufficient role in producing the empirical world. For example, it is highly implausible that the production of appearances that precede my birth would be within my

---

control.<sup>31</sup> This is so even supposing that as a thing in itself I am not in time. For even if I as a thing in itself am not in time, it is not plausible that I have ever done anything as a thing in itself that has genuinely affected anything that happened before my birth. Here we encounter a disanalogy to the lucid dreaming case – while I could freely cause the dream-content of the Martians causally determining me to steal the Mona Lisa, it is not plausible that I as a noumenal agent could freely cause phenomenal events that precede my birth.

In one text in the *Critique of Practical Reason* Kant appears to claim that the noumenal subject produces everything in the past that determines her free actions:

But the same subject, which, on the other hand, is conscious also of his own existence as a thing-in-itself, also views his existence so far as it does not stand under temporal conditions, and to himself as determinable only by laws which he gives to himself through reason. In this existence nothing is antecedent to the determination of his will; every action and, in general, every changing determination of this existence according to inner sense, even the entire history of his existence as a sensuous being, is seen in the consciousness of his intelligible existence only as a consequence, not as a determining ground of his causality as a noumenon. From this point of view, a rational being can rightly say of any unlawful action which he has done that he could have left it undone, even if as appearance it was sufficiently determined in the past and thus far was inescapably necessary. For this action and everything in the past which determined it belong to a single phenomenon of his character, *which he himself creates*, and according

---

<sup>31</sup> Jonathan Bennett raises similar objections in his “Kant’s Theory of Freedom,” in *Self and Nature in Kant’s Philosophy* (Ithaca: Cornell University Press, 1984), pp. 102-12, at pp. 102-3.

to which he imputes to himself as a cause independent of all sensibility the causality of that appearance. (*Ak* V 97-8, from Lewis White Beck's translation of Kant's *Critique of Practical Reason* (Indianapolis: Bobbs-Merrill, 1956, emphasis mine)<sup>32</sup>

However, this sort of atemporalist line is at best insignificantly more credible than an overt contradiction.<sup>33</sup> Moreover, it is far from clear that this is Kant's considered position, rather than a

---

<sup>32</sup> Allen Wood cites this passage in "Kant's Compatibilism," pp. 90-1. Ralph Walker points out problems for this conception in *Kant* (London: Routledge and Kegan Paul, 1978), pp. 148-9.

<sup>33</sup> Even if one would be willing to entertain seriously the possibility of our being atemporal agents, there is a problem that faces the particular story Kant wants to tell about how it works, and indeed any account of human atemporal agency. Kant's idea, discussed in *Religion Within the Boundaries of Mere Reason*, and presented by Wood ("Kant's Compatibilism," pp. 89-99) and Allison, (*Kant's Theory of Freedom*, pp. 47-53, 136-45), is that the fundamental atemporal decision we make is whether or not to subordinate the moral law to self-interest as the "supreme ground of our maxim" (*Rel*, *Ak* VI 32-44). Kant proposes that original sin consists in each of us indeed making the decision to subordinate the moral law to self-interest, and that this decision sets our moral character in the phenomenal world. However, it is important to Kant that each of us is also able to act for the sake of the moral law when self-interest motivates us to do otherwise: "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..." (*Rel*, *Ak* VI 45). This implies that we are capable of a moral conversion from original sin: "If by a single and unalterable decision a human being reverses (*umkehrt*) the supreme ground of his maxims by which he was an evil human being... he is to that extent, by principle and attitude of mind, a subject receptive to the good..." (*Rel*, *Ak* VI 48). But the very notion of reversal implies succession, which seems to be an essentially temporal notion. Kant also refers to this reversal as a decision for *restoring* (*widerherstellung*) the correct order in the supreme ground of maxims succeeds the decision for the incorrect order (e.g., at *Rel*, *Ak* VI, 44), and the notion of restoration also seems essentially temporal. Now the most prominent philosophical model for an atemporal decision is theological. On the traditional model, divine decision, which appears complex in time, is really one single atemporal act. However, it does not seem that a moral conversion from evil to good can be accounted for by a single atemporal decision. This won't be a worry in the divine case, since God never acts immorally. Perhaps we can conceive of an eternal decision to be morally evil for the first twenty years of one's temporal life, and morally good for the rest. But such an eternal decision does not amount to a genuine moral conversion, for that would require an actual change in one's moral disposition. Kant's own view does specify the possibility of a moral conversion, but it seems unavoidably to involve temporal succession, so it will not work as a model for atemporal decision. Moreover, it would seem that people can in fact undergo moral conversion, and since atemporal agency cannot account for that fact, it amounts to an implausible

view he experiments with in just one place.<sup>34</sup>

To my mind, a more credible sort of reconciliation can be provided by bringing in divine agency. 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>35</sup> On a version of

---

proposal. For further suggestions on how atemporal agency does not fit well with the moral life as we understand it, see Allen Wood, “Kant’s Compatibilism,” pp. 97-9.

<sup>34</sup> For example, it is not clear that this is his position in the *Critique of Pure Reason*. The best supporting passage for a view of this sort there is found at A548/B576:

*die Vernunft... folgt nicht der Ordnung der Dinge, so wie sie sich in der Erscheinung darstellen, sondern macht sich mit völliger Spontanität eine eigene Ordnung nach Ideen,* (Reason... does not follow the order of things as they are presented in appearance (*Erscheinung*), but with complete spontaneity it makes its own order according to ideas,

upon which follows the crucial phrase:

*in die sie die empirischen Bedingungen hinein paßt*

The passage then continues

*und nach denen sie sogar Handlungen für notwendig erklärt, die doch nicht geschehen sind und vielleicht nicht geschehen werden...* (and according to which it even declares actions to be required, that *have not occurred* and perhaps will not occur).

Kemp-Smith translates the emphasized passage as “to which it adapts the empirical conditions,” and Wood and Guyer as “to which it fits the empirical conditions.” The ‘to which it adapts’ makes it sound as if Kant is proposing that the (freedom-involving) order of reason is put in place, whereupon reason manipulates the empirical conditions to fit that order. But the ‘to which it adapts’ is an *interpretation* of the German, which more literally reads: “into which it fits the empirical conditions.” This more literal reading is certainly compatible with the last interpretation, but the passage itself does not even entail that by means of free choice we manipulate empirical conditions. Moreover, although this text suggests that Kant reconciles the empirical conditions with moral ‘ought’ principles, it does not state that by means of reason an agent freely produces empirical conditions that precede her birth, much less that it freely causes the entire empirical world. How reason fits the empirical conditions into the order of its own is here left unstated.

<sup>35</sup> Luis de Molina, *Liberi Arbitrii cum Gratiae Donis, Divina Praescientia, Providentia, Praedestinatione et Reprobatione Concordia* (1595); tr. (of Part IV) A.J. Freddoso, *On Divine Foreknowledge: Part IV of the Concordia* (Ithaca: Cornell University Press, 1988). For an

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 supposing the credibility of Kantian idealism more generally, the Kantian version of Molinism would be no less credible. True, for some this Kantian version of Molinism will not be credible because of the idealistic or the theistic beliefs that it presupposes, or because they find the objections to Molinism decisive.<sup>36</sup> But again, for the purposes of this discussion I don't want to set the standard of credibility too high, and it suffices that idealism and Molinism are not uncommonly accepted by people who have considered these views ably and seriously.<sup>37</sup>

---

excellent exposition and defense of Molina's position, see Thomas Flint, *Divine Providence* (Ithaca: Cornell University Press, 1998).

<sup>36</sup> A classic source of the "grounding" objection is Robert Adams, "Middle Knowledge and the Problem of Evil," *American Philosophical Quarterly* 14 (1977), pp. 109-17. For a rebuttal, see Thomas Flint's *Divine Providence*, pp. 121-37.

<sup>37</sup> Eric Watkins recently has advocated the following resolution to the problem at hand, which is in its broad contours similar to the suggested Molinist solution, although it does not specify the mechanism whereby the phenomenal laws are fixed by the noumenal choices, which for Molinism is divine agency:

Kant's solution to the modal conflict between free will and determinism depends on the possibility that the laws of nature are grounded in things in themselves in general and our free noumenal choices in particular. Kant makes this possibility more concrete in two ways. First, Kant holds that the laws of nature depend on the *natures* of things. Second, in the specific case of human beings, Kant wants to suggest – in line with our common sense views – that we can choose our own characters or natures. These two points, taken together, allow Kant to say, contra van Inwagen, that the laws of nature could in fact be up to us. The contingency of free will is thus compatible with the necessity of the laws of nature because our free will could choose (at least some of) the natures upon which the laws of nature, along with their necessity, are based. (*Kant and the Metaphysics of*

On the rival two-world position, no entity in the empirical world is identical to the transcendently free agent, and the issue of a single agent having incompatible properties does not arise. Rather, a concern for this view is the overdetermination issue – by one strand of its causal history the empirical action has a sufficient cause in a transcendently free subject, while by another strand this same action has a sufficient cause in a deterministic series of events that traces back to a time before the agent was born. The dream analogy, however, provides a model as to how it might be that transcendently free choices would dovetail in their effects in the empirical world precisely with how these effects are causally determined by empirical preceding conditions. To adapt the lucid dreaming example to the two-world view, I might freely will that I dream that my dream-self, now not identical to me, is deterministically manipulated by sophisticated Martian neuroscientists to steal the Mona Lisa. In these circumstances, I freely cause the dream-action of stealing the Mona Lisa, while the dream-self is causally determined to carry out the dream-action by the neuroscientists. This story makes it understandable that there might be overdetermination of the sort that Kant's theory involves. The credibility of this theory will here also depend on the details of the idealistic story Kant would need to tell, and the Kantian adaptation of Molinism seems to me to provide the best version.

5. Even though according to Kant we cannot establish on theoretical grounds that we are transcendently free, he wants to argue that it is nevertheless legitimate for us to believe that we are free in this sense. The grounds for legitimacy are practical – we have reasons that derive from morality in particular for believing that we are transcendently free. Kant presents two such

---

*Causality*, pp. 344-5).

Watkins makes a good case for the claim that this is Kant's view of the matter.

reasons. The first is that for us moral principles 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).

However, Kant contends, “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.

One idea Kant might be expressing is that there is nothing in the natural world that can ground normativity in general. But his specific concern here is with moral and perhaps prudential ‘ought’ judgments, and how transcendental freedom might help provide the requisite account. Later, in his *Religion Within the Boundaries of Mere Reason*, Kant explicitly defends an ‘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>38</sup> Thus it seems reasonable to interpret Kant as supposing here (at A547/B575) that if ‘ought’ principles are true or hold for us, it must in general be the case that we are able to act in accord with them. The following moral ‘ought implies can’ principle is attractive: If one ought to do something, then it must be the case that one can do it.<sup>39</sup> Accordingly, if because one is causally determined one can never do otherwise, then it would be false that one ever ought to do otherwise. Moreover, if it is never true that one ought to do otherwise, what would be the point of a system of moral ‘ought’s? 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>40</sup> So, by this argument, it would seem that ‘ought’ principles cannot

---

<sup>38</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>39</sup> For a sensitive discussion of these issues, see Walter Sinnott-Armstrong, “‘Ought’ Conversationally Implies ‘Can,’” *The Philosophical Review* XCIII (1984), pp. 249-261, and “‘Ought To Have’ and ‘Could Have,’” *Analysis* 45 (1985), pp. 44-48. See C. D. Broad, “Determinism, Indeterminism, and Libertarianism,” in his *Ethics and the History of Philosophy* (London: Routledge and Kegan Paul, 1952), pp. 195-217, reprinted in *Determinism, Free Will, and Moral Responsibility*, Gerald Dworkin, ed. (Englewood Cliffs: Prentice-Hall, 1970) pp. 149-71, for a defense of the claim that determinism undermines central judgments of moral obligation. Broad also describes types of ‘ought’ judgments that he thinks are not threatened by determinism, for example:

When we say that a man ought not to cheat at cards we often mean to assert two things. (a) That the average decent man does not do this, and that anyone who does falls in this respect below the average. And (b) that a man who does this either has a very low ideal of human nature or a very weak and unstable desire to approximate the ideal which he has. So in this further respect, he falls below the average. (pp. 159-60)

I discuss these matters in more detail in *Living Without Free Will*, pp. 141-8.

<sup>40</sup> See Ishtiyaque Haji, *Moral Appraisability* (New York: Oxford University Press, 1998), and *Deontic Morality and Control* (Cambridge: Cambridge University Press, 2002) for a more

hold or be true if we had no capacity for action that was not subject to natural causal determinism. The idea would then be that given that we have a good practical reason to preserve the ‘ought’ judgments, we thereby have a good practical reason to believe that we are transcendently free.

But how strong is this practical reason? An initial problem is that what would be required for moral principles to be true for us or hold for us is not the belief that we are transcendently free, but rather our actually being transcendently 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 Within the Boundaries of Mere Reason* 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.<sup>41</sup> 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 or hold for us, rather than the truth of the moral principles or their holding for us, that Kant aims to secure. Given this supposition, the belief that we are transcendently 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

---

thoroughly developed argument for this sort of claim.

<sup>41</sup> *Rel*, Ak VI 6-8n. See also Robert Adams, “Moral Arguments for Theistic Belief,” reprinted in his *The Virtue of Faith* (Oxford: Oxford University Press, 1990).

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' 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 have done otherwise. Frequently, it is significantly probable that expressing a moral 'ought' judgment will causally influence the selection of options for action, and thus there is a good moral 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 Haji, 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>42</sup> Haji presupposes, as C. D. Broad also contended,<sup>43</sup> that 'ought' judgments have various distinct roles, and that these roles

---

<sup>42</sup> Haji writes: "For the argument for the incompatibilist of determinism and deontic morality is not in any way concerned with the action-guiding function of ought judgments" *Deontic Morality and Control*, p. 77.

<sup>43</sup> C. D. Broad, "Determinism, Indeterminism, Libertarianism," in *Determinism, Free Will, and Moral Responsibility*, in Dworkin, at p. 169.

have different truth or assertability conditions. The truth of the action-guiding variety of 'ought' judgment is not affected by determinism and by agents' inability to have done otherwise, even if 'ought' judgments in other roles would then be undermined.

Secondly, even if moral 'ought' judgments do turn out not true or do not hold for us because we 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 or hold for us. 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 to do x' and 'it is morally bad to do y' still could be true or hold for us. 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 all of the roles that one might think 'ought' judgments have in guiding action – in moral encouragement and admonition -- and in moral evaluation as well. If this is so, and since the truth of these alternative principles does not appear to require transcendental freedom, the practical consideration for believing that we are transcendentally free that Kant adduces here is not clearly very strong.<sup>44</sup>

---

<sup>44</sup> One might argue that one capacity central to Kant's ethical framework does require transcendental freedom – viz. autonomy. However, the core feature of Kantian autonomy, positive freedom, does not really presuppose transcendental freedom (despite what Kant says at A534/B562). To have positive freedom, by Thomas Hill's characterization, is to have a capacity to commit oneself to certain principles of conduct as rationally binding, principles that are not adopted to satisfy any contingent desires, but necessarily imposed on oneself as a rational agent

6. 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 ordinary sense of morality does. When people perform actions contrary to the moral law, we typically judge them blameworthy. 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

---

("The Kantian Conception of Autonomy," in Thomas Hill, *Dignity and Practical Reason* (Ithaca: Cornell University Press, 1993, pp. 76-96). Now if the previous argument holds, and certain kinds of ‘ought’ judgments and principles of right and wrong generally do not require transcendental freedom, no further obstacle remains to having positive freedom without transcendental freedom. For there is now nothing about the commitment in question that precludes an agent’s being causally determined to make it by factors beyond her control.

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 preceding state, as though with that act the agent had started a series of consequences entirely from himself. (A554-5/B582-3)<sup>45</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 recall 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 superficial 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 toward 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 last murderer remaining in prison in Kant's imagined island society that is about to dissolve itself (in *The Metaphysics of Morals*). Kant strenuously

---

<sup>45</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)

advocates that he should be executed, just because of the crime he has committed; that is, for reasons of retributive desert alone. (Ak VI 331-3) But imagine the offender protesting that he was determined by natural causes to act as he did. Would the following reply count as morally acceptable? “Although we have no evidence of your transcendental freedom, and although we cannot show that such a power of agency is 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 deserving of punishment.” It would not. Holding an offender blameworthy, expressing one's anger towards him, and depriving him of life or liberty all tend to be harmful to the offender. In general, if one aims to harm another, then one's justification must meet a high epistemic or theoretical standard – much higher than the standard of consistency that Kant advocates, and higher than a standard that merely yields credibility, especially credibility only for some people. If it is significantly probable that one's justification for the harming another is unsound, then, *prima facie* that behavior is seriously wrong and one must refrain from engaging in it. If one's justification for harmful behavior depended on the claim that we are transcendently free, and we have little or no evidence for this claim, and the story we need to tell to reconcile transcendental freedom with our best empirical theories is barely credible, then that justification is inadequate.

Wood likens Kant's attempt to justify a belief in transcendental freedom to a defense attorney's attempt to show that his client's innocence cannot be ruled out.<sup>46</sup> Kant may indeed have established that our being transcendently free cannot conclusively be ruled out. However, in a large range of cases, such as that of the malicious liar, or the last murderer on death row in

---

<sup>46</sup> Allen Wood, “Kant's Compatibilism,” p. 83.

the island society, the guidance that belief in our transcendental freedom would provide is more aptly described as on the side of *prosecuting* attorney. The epistemic standard that the prosecuting attorney must meet is not merely that his claims cannot conclusively be ruled out. Thus there is a moral reason why a belief in transcendental freedom requires much stronger evidential or theoretical grounding than it has on Kant's account.

Accordingly, this second reason for the belief that we are transcendently free, like the first, does not seem impressively strong. True, there may be more to these reasons than my analysis has brought out. Moreover, there are further practical reasons to believe that we are free in this sense that we haven't considered, for example, reasons that derive from the importance of moral worth and personal achievement. However, I think we can conclude that some of the practical reasons for believing that we are transcendently free that are often thought to carry significant weight do not obviously do so, and that as a result we should not assume, without careful assessment, that believing that we are free in this sense is practically warranted.

7. A further response to Kant's "malicious lie" passage is that he is mistaken to assume that the practice of holding people blameworthy requires the grounding in transcendental freedom he advocates for it, for the reason that the practice retains its legitimacy regardless of whether our actions have causal histories that are exhausted by deterministic natural causes. P. F. Strawson famously developed this response, and many have since found it attractive.<sup>47</sup> A feature of Strawson's account is that nothing internal to the practice of holding people morally responsible demands the grounding of that practice in transcendental freedom. The move to transcendental

---

<sup>47</sup> P. F. Strawson, "Freedom and Resentment," *Proceedings of the British Academy* 48 (1962), pp. 1-25.

freedom is a response to an illegitimate request for justification of holding people morally responsible, and the illegitimacy of this demand can be traced to the fact that it is external to the practice.

On this issue, I believe that Kant is right and Strawson is mistaken. One of the reasons is that features that are, after all, internal to the practice itself generate the demand for justification that Kant answers with his postulation of transcendental freedom. As Strawson notes, the practice includes provisions for exemption from moral responsibility on various sorts of grounds. For example, we regard someone who is brainwashed into committing crimes as exempt from moral responsibility. But this sort of consideration can naturally be extended to cases in which agents are covertly controlled by other agents, or by machines, and causally determined to commit crimes. Indeed, even if they meet the various proposed compatibilist conditions on moral responsibility, our intuitions – continuous with our actual practice of holding people morally responsible – exempt them from moral responsibility. Furthermore, it is a feature of our practice of holding people morally responsible that if no relevant moral difference is to be found between agents in two situations, then if one agent is legitimately exempted from moral responsibility, so is the other. And, as I have argued, there is no relevant moral difference between agents in these covert control cases and agents whose actions have causal histories that are exhaustively deterministic.<sup>48</sup> So it is the practice itself, in particular central rules internal to the practice, that renders Kant's verdict that agents are not morally responsible for actions whose causal histories are exhausted by deterministic natural causes.

Advocates of Strawson's view also suggest that Kant's contention that exhaustively

---

<sup>48</sup> See my *Living Without Free Will*, pp. 111-6, 123.

deterministic causal histories of actions would undermine moral responsibility is merely a skeptical worry, and just as for our practice of induction, the inability to answer the skeptic does not rationally constrain us to give up the practice and the beliefs it involves.<sup>49</sup> Indeed, few would deny that we can legitimately retain the practice of induction and its attendant beliefs even if we cannot answer Hume's skeptical argument. However, there are important differences between the practice of induction and the practice of holding people morally responsible. First, we cannot as a matter of psychological fact give up the practice of induction – not even a part of it. Second, we can plausibly conceive of Hume's challenge as external to the practice – as not generated by any features of the practice itself -- while it remains internally unchallenged. But the analogous claims are not true of the practice of holding people morally responsible. We can and in fact have given up parts of it. For example, over the past several centuries many have given up judgments of moral responsibility in the case of criminals who are mentally ill. In addition, as I have argued elsewhere, it may be that we can give up all of it without significant practical loss. Moreover, as is indicated by developments in how we regard such criminals, the practice has been opposed from the inside by local deterministic considerations. To cite a recent example of this sort of challenge, a study from New Zealand released in 2002 indicates that 85 percent of boys who have a weakened version of a gene that controls production of an enzyme called monoamine oxidase A -- which breaks down key neurotransmitters linked with mood, aggression and pleasure -- and who were abused turned to criminal or antisocial behavior (30% of the study group has the weakened version of the gene).<sup>50</sup> For many people, this sort of discovery occasions doubt about

---

<sup>49</sup> Thanks to Jonathan Adler for raising this issue.

<sup>50</sup> "Role of genotype in the cycle of violence in maltreated children," A. Caspi, J. McClay, T. E. Moffitt, J. Mill, J. Martin, I. W. Craig, A. Taylor, and R. Poulton, *Science*, Vol. 297, No.

the moral responsibility of criminals who fit this description. And in addition, as we have just seen, global or universal determinism can be viewed as posing an internal challenge to the practice as a whole.

So I don't believe that Kant is mistaken in his sense that the practice of holding people morally responsible requires the sort of transcendental grounding that he proposes. This transcendental grounding has often been regarded as paradigmatic of a response to an external demand for the justification of a practice, but instead it is plausibly a response to an internal demand. However, even if Kant's theory of transcendental freedom was a response to an external demand, it would not thereby be illegitimate. Few would be willing to agree that external demands for justification are illegitimate in the case of, say, religious or racist practices. In fact, note that in the racism case we hold that unanswered challenges plausibly external to the practice – for example, a challenge from an absence of difference in skills and abilities – will serve to undermine the legitimacy of beliefs that form part of it, no matter how deeply ingrained the practice is, and no matter how difficult it is for people to renounce the beliefs, attitudes, and behavior that constitutes it. Perhaps the practice of holding people morally responsible is

---

5582, August 2002, 851-4. Here is a summary from *Crime Times* 8 (2002), pp 1-2: Avshalom Caspi and colleagues analyzed data from 442 New Zealand male adults involved in a long-term study. The researchers identified 154 subjects who were abused or maltreated as children, including 33 who were severely abused. The researchers then evaluated the influence of a particular gene on the abused children's outcomes as adults. A "low activity" variant of this gene which affects levels of monoamine oxidase A (MAOA), an enzyme that metabolizes the brain chemicals serotonin, dopamine, and norepinephrine has previously been linked to abnormal aggression. Caspi et al. discovered that 85 percent of severely abused subjects with the low-activity variant of the MAOA gene developed some form of antisocial behavior. In contrast, study participants with the high-activity variant only rarely exhibited aggressive or criminal behavior in adulthood even if they had been severely abused as children. "Although individuals having the combination of low-activity MAOA genotype and maltreatment were only 12 percent of the male birth cohort," the researchers say, "they accounted for 44 percent of the cohort's violent convictions."

similarly vulnerable to Kant's challenge even if it is appropriately regarded as external.

8. 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, and whether we have transcendental freedom cannot be established on evidence available to us. In addition, Kant provides a consistent conception according to which we have freedom of this sort; and with some assistance from Molina, this conception meets a significant standard of credibility. At the same time, I have argued that the practical reasons that Kant adduces for belief in transcendental freedom are subject to serious challenge. But, again, there may be more to these reasons than I have seen, and there are other reasons for believing in transcendental freedom that must also be considered. Moreover, the credibility and commitments of alternative views of free will and action must be weighed against Kant's. In the last analysis, it is fair to say that in its general contours Kant's position on free will remains a significant and attractive option.<sup>51</sup>

---

<sup>51</sup> This paper has benefitted from discussions at the CUNY Graduate Center; the University of Massachusetts, Amherst; the University of California, San Diego; and the University of Colorado, Boulder. Special thanks to Richard Arneson, Philip Bricker, David Christensen, Hilary Kornblith, Dana Nelkin, Sam Rickless, Eric Watkins, and the referees for *Philosophy and Phenomenological Research*.