



Contents lists available at ScienceDirect

Stud. Hist. Phil. Sci.

journal homepage: [www.elsevier.com/locate/shpsa](http://www.elsevier.com/locate/shpsa)



# Empirical psychology, common sense, and Kant's empirical markers for moral responsibility

Patrick Frierson

Whitman College, 345 Boyer Avenue, Walla Walla, WA 99362, USA

## ARTICLE INFO

### Keywords:

Immanuel Kant  
Moral responsibility  
Freedom  
Empirical psychology  
Transcendental idealism  
Reason  
Common sense  
Anthropology

## ABSTRACT

This paper explains the empirical markers by which Kant thinks that one can identify moral responsibility. After explaining the problem of discerning such markers within a Kantian framework, I briefly explain Kant's empirical psychology. I then argue that Kant's empirical markers for moral responsibility—linked to higher faculties of cognition—are not *sufficient* conditions for moral responsibility, primarily because they are empirical characteristics subject to natural laws. Next, I argue that these markers are not *necessary* conditions of moral responsibility. Given Kant's transcendental idealism, even an entity that lacks these markers *could* be free and morally responsible, although as a matter of fact Kant thinks that none are. Given that they are neither necessary nor sufficient conditions, I discuss the status of Kant's claim that higher faculties are empirical markers of moral responsibility. Drawing on connections between Kant's ethical theory and 'common rational cognition' (4:393), I suggest that Kant's theory of empirical markers can be traced to ordinary common sense beliefs about responsibility. This suggestion helps explain both why empirical markers are important and what the limits of empirical psychology are within Kant's account of moral responsibility.

© 2008 Published by Elsevier Ltd.

When citing this paper, please use the full journal title *Studies in History and Philosophy of Science*

## 1. Introduction

In the *Critique of practical reason*, Kant writes that 'the moral law commands compliance from everyone' (5:36, see also 29:603).<sup>1</sup> In the *Anthropology*, he reiterates that 'the law of duty ... is present in

everyone' (7:214). But in lectures on empirical psychology, Kant claims that 'in some cases ... [a human being] has no power of free choice, e.g., in the most tender childhood, or when he is insane, and in deep sadness, which is however a kind of insanity' (28:255, from *Metaphysics L1*). Given that for Kant free choice is necessary for mor-

E-mail address: [frierspr@whitman.edu](mailto:frierspr@whitman.edu)

<sup>1</sup> Throughout, references to Kant are given using the Academy Edition pagination (Kant, 1900–). For the first *Critique*, references are to the A and B editions. Where available, I have used translations from *The Cambridge edition of the works of Immanuel Kant in translation* (Kant, 1995–). Translations of the lectures on anthropology (Ak. 25) are my own. References to Kant's lectures are to specific sets of lecture notes. In the text, I give the Academy Edition pagination for these lectures, but this pagination corresponds to the following customary titles of the lecture notes (generally based on the student who transcribed the notes), listed here in the order of their page numbers in the Academy Edition: *Lectures on metaphysics*: Herder (28:1–166, 869–962; dated 1762–1764); *L1* (28:167–350; dated mid-1770s); *Volckmann* (28:351–459, dated 1784–1785); *von Schön I* (28:461–524, probably dated 1780s); *L2* (28:531–610, probably dated 1790–1791); *Dohna* (28:615–702, dated 1792–1793); *K2* (28:705–816, dated early 1790s); *Mrongovius* (29:747–940, dated 1782–1783); *Vigilantius* (29:943–1040, dated 1794–1795).

*Lectures on ethics*: Herder (27:3–89, dated 1762–1764); *Powalski* (27:93–235, probably dated 1782–1783); *Collins* (27:243–471, dated 1784–1785); *Vigilantius* (27:479–732, dated 1793); *Mrongovius* (29:597–744, dated 1785).

*Lectures on anthropology*: *Collins* (25:1–238, dated 1772–1773); *Parow* (25:239–464, dated 1772–1773); *Friedländer MS 399/400* (25:465–728, dated 1775–1776); *Pillau* (25:729–848, dated 1777–1778); *Menschenkunde* (25:849–1204, probably dated 1781–1782); *Mrongovius* (25:1205–1431, dated 1784–1785); *Busolt* (25:1431–1532, probably dated 1788–1789).

al responsibility, this implies a scientific basis for claiming that children and the insane are not morally responsible.<sup>2</sup> In his *Anthropology*, Kant even allows that when ‘someone has intentionally caused harm’, the question can still arise ‘whether he is guilty of it and to what extent, so that the first thing to be determined is whether or not he was mad at the time’ (7:213).<sup>3</sup> And in another set of lectures, Kant insists that when someone ‘pushes another into the water ... and that person drowns’, there is still a question about whether such a person is morally responsible for this deed. The ‘push’ might, for instance, have been simply the consequence of ‘dizziness’ or some other ‘cause [that] was merely physical and a matter of natural necessity’ (27:559, *Ethics Vigilantius*). Moreover, in these discussions of the limits of moral responsibility, Kant seems to claim that the question of moral responsibility is an *empirical* one, one that ‘is purely psychological’ (7:214). Ultimately, however, while Kant gives a detailed account of the empirical markers for moral responsibility—markers that it is the business of empirical psychology to study—this account rests on a set of more basic commonsense moral intuitions about when one is morally responsible. For Kant, empirical psychology ends up playing an important but fundamentally secondary role in ascriptions of moral responsibility.

The claim that almost all human beings are morally responsible but that some human beings (such as children) or human beings at certain times (such as when mad or dizzy) are not morally responsible seems fairly sensible. And the claim that empirical psychologists are best qualified to judge whether or not a person is morally responsible is at least plausible and is widely accepted in jurisprudential practice. But the attempt to carve out ground for these fairly sensible positions raises an important problem for Kant. Kant’s first *Critique* argues that although every event in nature is causally determined, it is nonetheless possible that the ultimate grounds of at least some events lie in free agents. In the first *Critique*, Kant’s defense of freedom is extremely limited. As he says, ‘we have not been trying to establish the reality of freedom ... [nor even] the possibility of freedom ... [but only] that nature at least does not conflict with causality through freedom’ (A558/B586). In the *Groundwork* and the second *Critique*, Kant goes further, seeking to show the reality<sup>4</sup> of freedom, at least in the case of human agents. The argument of the *Groundwork* argues from the consciousness of the idea of freedom to participation in an intelligible realm and thus to actual freedom (4:452).<sup>5</sup> By the time of the second *Critique*, Kant seems to have rejected this argument in favor of a more straightforward regressive<sup>6</sup> proof of freedom as the condition of the possibility of moral responsibility. As Kant puts it there, one ‘judges that he can do something because he is aware that he ought to do it, and cognizes freedom within him, which, without the moral law, would have remained unknown to him’ (5:30).

Kant’s shift to this regressive argument in the second *Critique* has been criticized for not dealing adequately with the skeptical concerns that prompted his earlier attempts to find a non-moral argument for freedom.<sup>7</sup> But the shift in argumentative strategy also raises a specific problem for identifying moral responsibility. Kant seems to assume in the second *Critique* that the ascription

of moral responsibility to an entity is trivial, at least absent any skeptical doubts. But often it is not. As Kant points out, in the cases of children, the insane, and even those in ‘deep sadness’, it becomes unclear where to draw the line. One may extend these concerns about moral responsibility to other human cases, and even to animals. On what grounds, for instance, do we justify holding most human beings morally responsible and not chimpanzees or dolphins?<sup>8</sup> The argument of the first *Critique*, showing that natural necessity does not conflict with freedom, works just as well for these animals as for humans. One cannot use Kant’s strategy in his early ethics lectures of distinguishing cases of responsibility from those in which the cause is ‘a matter of natural necessity’ (27:559), because according to Kant’s transcendental idealism every human deed fits into a series of events that is governed by natural necessity.<sup>9</sup> The universality of natural necessity seems to cut off the possibility of the most obvious sorts of empirical-scientific tests for ‘freedom’. And one cannot—at least by the time Kant rejects the arguments of the *Groundwork*—argue from a person’s (or animal’s) transcendental freedom to their moral responsibility because there is no way to prove that any being is transcendently free except from the conditions of possibility of moral responsibility. So how can Kant distinguish between those who are morally responsible and those who are not? And even in cases of moral agents, how can one distinguish acts or dispositions for which one is morally responsible from the sadness and madness that absolves one of guilt?

In answering these questions, it is important to avoid the temptation of the *Groundwork*. The Kantian should not seek to reason from a particular empirical psychology to a non-moral proof of freedom as a ground of moral responsibility. Kant does acknowledge that the moral law, and with it moral responsibility, ‘would be analytic if the [transcendental] freedom of the will were presupposed’ (5:31, see also 4:447). But this claim does not provide a way to get from empirical psychology to moral responsibility because this sort of freedom cannot in principle be experienced by human beings. By the second *Critique*, Kant gives up on any attempt to provide a proof of freedom independent of human moral responsibility, arguing instead that one is immediately conscious of one’s responsibility to obey the moral law.

Consciousness of this fundamental law may be called a fact of reason because one cannot reason it out from antecedent data of reason, for example, from consciousness of freedom ... and because it instead forces itself upon us as a synthetic a priori proposition that is not based on any intuition. (5:31).

Given this ‘fact of reason’, one can show that human beings are transcendently free. But there is no proof of this fact itself. Even in the first *Critique*, Kant recognizes that although ‘the question of the possibility of freedom does indeed assail psychology’, yet ‘since it rests merely on dialectical arguments of pure reason, its solution must be solely the business of transcendental philosophy’ (A535/B563). And Kant is even clearer in a metaphysics lecture from the early 1790s:

<sup>2</sup> I use the term ‘scientific’ here loosely. For discussions of the scientific status of psychology in Kant, see Frierson (2005a); Hatfield (1992); Sturm (2001).

<sup>3</sup> In the first *Critique*, Kant says something similar: ‘the real morality of actions ... remains hidden ... [because] how much of [our actions] is to be ascribed to mere nature and innocent defects ... this no one can discover’ (A551/B579 n.).

<sup>4</sup> As Kant emphasizes, this ‘reality’ is ‘only for practical purposes’ (5: 133). Kant does not claim to prove that freedom is real in an empirical or speculative sense.

<sup>5</sup> This argument has been widely discussed. Cf. Allison (1990), pp. 214–229; Ameriks (1981); Beck (1960), pp. 109–125.

<sup>6</sup> I use the term ‘regressive’ here in reference to Karl Ameriks’s helpful account of the regressive structure of Kant’s arguments in the first *Critique*. See Ameriks (2003).

<sup>7</sup> See *ibid.*

<sup>8</sup> Within Kantian ethics this has important implication for the scope of moral regard as well, because the sole criterion for moral regard is the capacity for having a good will. As Kant explains in the *Metaphysics of morals*, ‘a person is a subject whose actions can be imputed to him’ (6:223; original emphasis). Thus if animals can be held morally responsible, then they may be considered ‘persons’, and human beings may have direct obligations to them. However, Patrick Kain has recently drawn attention to the fact that for Kant, the mere capacity for moral responsibility, even if this is never actualized, may be sufficient for moral regard. Cf. Kain (n.d.).

<sup>9</sup> This claim is defended in detail in Frierson (2005a).

Freedom cannot be proven psychologically, but rather morally. Through morality I consider a human being not as a natural being, as object of the senses, but rather as intelligence, as object of reason. If I wanted to prove freedom psychologically, then I would have to consider a human being according to his nature, i.e., as a natural being, and as such he is not free. (*Metaphysics* K2, 28:773; see also *Metaphysics* Dohna 28:682, ca. 1792–1793)

Because even inner experience is necessarily structured by the category of causation, there is no empirical psychology that can justify moral responsibility on its own. Moreover, given Kant's transcendental idealism, there is no empirical psychology that is incompatible with freedom. Paul Guyer has even suggested, on these grounds, that 'the subjective state of one's feelings' can, perhaps even directly, 'reflect the moral choices of one's will' (Guyer, 1993, p. 367).

Nonetheless, although Kant's transcendental idealism does not force him to adopt any particular empirical psychology, he does in fact develop a very specific account of human action at the empirical level, and this empirical picture is constructed in a way that highlights certain *empirical* features of human action that correlate with human freedom. In ethics lectures as late as the winter of 1793, for example, he describes a 'visible spontaneity' in certain actions that is 'an essential criterion of freedom' (27:505, *Ethics Vigilantius*). Kant quickly makes clear here that this visible spontaneity is *not* the transcendental freedom that is a *condition of the possibility* of moral obligation; immediately after referring to 'visible spontaneity' in human nature, Kant raises the possibility that actions proceeding from this spontaneity might be 'grounded, simultaneously, in the time preceding' such that 'unconditioned self-activity would not be present in it' (ibid.). Because 'it was this [unconditioned self-activity] that was demanded of man *qua* noumenon or intelligible being ... only as an intelligible being does he emerge completely from the world of the senses ... Freedom, therefore, cannot be made comprehensible' (ibid.). Still, Kant insists that this visible spontaneity is an important '*criterion*' (*Criterium*) of freedom. And in his anthropology Kant refers to character, an empirically recognizable capacity of a human being that is associated with one's visible spontaneity, as 'a mark [*Merkmal*] of a rational being' (25:1156) or even a 'distinguishing sign [*Unterscheidungszeichen*] of a rational being endowed with freedom' (7:285).<sup>10</sup> Visible spontaneity—the sort that could be part of an empirical psychology—is neither a necessary nor a sufficient condition for transcendental freedom, but it can still be an important empirical criterion, sign, or marker of it.

In his *Groundwork*, and in greater detail in his second *Critique*, Kant gives even more detailed accounts of empirical correlates to moral freedom, describing what the moral law empirically 'effects in the mind insofar as it is an incentive' (5:72, see also 4:400). In his *Metaphysics of morals*, when Kant explains 'the relation of the human mind to moral laws' (6:211), he draws on his *empirical* account of human psychology to explain where to situate moral motives in human psychology; and in his *Critique of practical reason* he explains that 'the explication [of the concepts of the faculty of desire and the feeling of pleasure] as

given in psychology could reasonably be presupposed' in these works (5:9 n.). And in his *Anthropology*, when he raises the possibility that a person might not be morally responsible for her actions, Kant insists that this question is 'purely psychological', to be solved by determining 'whether the accused was in possession of his *natural* powers of understanding and judgment' (7:213–214; my emphasis). Thus although Kant insists that 'by empirical psychology ... we should know ourselves merely in the world of sense ... [and] therefore morality is the sole means of obtaining consciousness of our freedom' (27:506, *Ethics Vigilantius*), he nonetheless holds that empirical psychology provides empirical *markers* that can be used to distinguish cases in which (transcendental) freedom is present from those in which it is not.

The first task of the rest of this paper will be to explain briefly Kant's empirical psychology insofar as this bears on moral responsibility. I then argue that Kant's empirical markers for moral responsibility are not sufficient conditions for moral responsibility, primarily on the grounds that they are empirical characteristics that are subject to natural laws. This causal determination, combined with Kant's insistence in the second *Critique* that transcendental freedom is a condition of the possibility of moral responsibility, shows that these criteria do not constitute a proof of moral responsibility. In this context, I describe two hypothetical situations within which one could meet all of Kant's empirical criteria and still fail to be morally responsible. Next, I argue that these empirical markers are not necessary conditions of moral responsibility. Given Kant's transcendental idealism, even an entity that lacks these empirical markers *could* be free and thus morally responsible, although as a matter of fact Kant thinks that none are. Finally, I suggest that a reading of Kant's ethics that emphasizes its connections with 'common rational cognition' (4:393) can explain both why empirical markers are important and what the limits of empirical psychology are within Kant's account of moral responsibility.

## 2. Empirical psychology and moral markers

Kant's empirical psychology is organized around three basic faculties of the soul—cognition, feeling, and desire.<sup>11</sup> Of these, the faculty of desire is the most important for understanding the empirical markers of moral responsibility because 'all desires have a relation to activity and are the causality thereof' (25:1514). Within each of his three faculties, Kant distinguishes between several basic powers, grouping these into 'higher' and 'lower' faculties of cognition, feeling, and desire. With respect to cognition, the higher faculty includes reason, the understanding, and judgment; the lower faculty includes the senses and imagination. With respect to the faculty of desire, Kant distinguishes the higher and lower faculties on the basis of the faculty of cognition that causes the relevant desire. For Kant, every desire is caused by some cognition, but one can distinguish between desires with causes that '*lie ... in the understanding*' and those with causes that lie '*in the sensibility*' (29:1014). The former are 'motives' and belong to the higher faculty of desire; the latter are 'stimuli' and belong to the lower faculty (29:1015, *Metaphysics Vigilantius*).

<sup>10</sup> The importance of character is discussed in greater detail in Munzel (1999); Kuehn (2001); Jacobs (2003); Frierson (2003). The empirical nature of character is discussed in Frierson (2003, 2005a); Jacobs (2003).

<sup>11</sup> There is a further aspect of Kant's anthropology that I do not discuss here, though it has some relevance to Kant's overall picture of human nature and thereby of human agency. In his *Anthropology* and *Critique of judgment*, Kant adds to this general account of faculties of soul an analysis of their underlying 'predispositions'. For more on these predispositions, see Kain (n.d.); Frierson (2005a); Sloan (2002), pp. 229–253; Zammito, (2003).



For Kant, the 'visible spontaneity' that marks moral responsibility is associated with the higher faculty of desire.<sup>12</sup> As Kant explains, 'the concept of freedom rests on this: namely the faculty of a human being for determining oneself to action through motives' (29:1016). And because the distinguishing feature of the higher faculty of desire is its determination by the higher faculty of cognition, the higher faculty of cognition becomes an important distinguishing mark of those endowed with moral responsibility. Thus Kant says, 'Reason is the persisting condition of all voluntary actions under which the human being appears' (A553/B581), and Kant uses the 'understanding' to distinguish the 'power of free choice [*arbitrium liberum*]', which can occur only with human beings, from the *arbitrium brutum* of animals (28:588, *Metaphysics* L2).<sup>13</sup> Likewise in the *Anthropology*, Kant insists that courts must refer the question of whether a criminal should be held morally responsible to empirical psychologists<sup>14</sup> because this issue rests on 'the question of whether the accused at the time of his act was in full possession of his natural faculty of understanding and judgment' (7:213).

This 'natural faculty of understanding' does not involve anything specifically moral and need not even be purely rational. Although all 'higher' desires have 'grounds of determination ... [that] lie ... in the understanding' (29:1014), these desires can be 'either pure or affected' (29:1015, *Metaphysics* *Vigilantius*). As Kant explains,

The intellectual impelling cause is either purely intellectual without qualification [*simpliciter talis, mere intellectualis*], or in some respect [*secundum quid*]. When the impelling cause is represented by the pure understanding, it is purely intellectual, but if it rests on sensibility, and if merely the means for arriving at the end are presented by the understanding, then it is said to be in some respect [*secundum quid*]. (28:589, *Metaphysics* L2)

When a desire is impure but still associated with the higher faculty of desire, one acts on the basis of a principle of the understanding

that is directed towards fulfilling some lower desire. Whereas animals may act in a law-like way in pursuit of their ends, they are not (according to Kant) motivated by principles that link specific actions to their respective ends.<sup>15</sup> Human beings can act on such principles. And although such principles amount to merely hypothetical imperatives, they still relate to the higher faculty of desire. In that sense, even a capacity to act on hypothetical imperatives is a marker of moral responsibility.<sup>16</sup>

In the *Anthropology*, Kant seems to associate moral responsibility merely with the capacity for action from a higher faculty of desire, and he suggests that for practical purposes, this issue reduces to the question of whether someone has a 'natural faculty of understanding'. But sometimes Kant suggests that the capacity for one's faculty of desire to be determined by pure reason is also an important marker of moral responsibility. That is, one who is morally responsible will have at least the capacity for impelling causes that are purely intellectual.<sup>17</sup> In the *Metaphysics of morals*, for example, Kant shifts from his standard definition of the free *Willkühr* or *arbitrium liberum* as the faculty of desire that is determined by the understanding, broadly construed. There he claims that 'That choice which can be determined by pure reason [Kant's emphasis] is called free *Willkühr*' (6:213). By contrast, in a lecture probably given in 1790–1791, Kant explains that 'power of free *Willkühr* [*arbitrium liberum*] can occur only with human beings, who have understanding' where this understanding can be 'either pure or affected' (28:588 L2, cf. 28:677 *Dohna*). The shift to pure reason in the *Metaphysics of morals* might reflect Kant's concern with defining the 'faculties of the human mind' insofar as these relate 'to moral laws' (6:211). And this suggests that a capacity for action on the basis of pure reason might be an important criterion of moral responsibility in human agents.<sup>18</sup>

This capacity to be motivated by the moral law itself also seems to play a role in Kant's accounts of moral psychology in the *Groundwork* and the second *Critique*. When Kant discusses 'respect for the

<sup>12</sup> One important challenge with applying Kant's empirical psychology in the context of markers of moral responsibility is that throughout Kant's accounts of human psychology he uses the same terms to refer to both noumenal bases and phenomenal causes of human action. Of these, the noumenal bases are in fact necessary (and in some cases sufficient) conditions of moral responsibility, while the phenomenal causes are merely markers for that responsibility. For example, in Kant's practical philosophy, he discusses a free noumenal power of choice (*Willkühr*) combined with a pure practical reason (*Wille*) that legislates for that power of choice. And Kant makes clear throughout his moral philosophy that this (transcendental) freedom is, as he puts it in the second *Critique*, the *ratio essendi* of the moral law. In that sense, a free *Willkühr* is a necessary and perhaps a sufficient condition of moral responsibility. (Arguably, a noumenal *Wille* is also a necessary condition of moral responsibility, and one might imagine entities that have a noumenal *Willkühr* without a *Wille*. If this is possible, then a free *Willkühr* would be a necessary but not sufficient condition of moral responsibility.) But Kant also discusses the free *Willkühr* as simply a capacity of certain organisms—human beings—to have desires that are caused by certain sorts of cognitions. And in this context, neither 'freedom' nor the '*Willkühr*' are necessary or sufficient for moral responsibility; they are simply (as we will see) empirical markers. (Beck, 1987, has discussed this ambiguity, (and others), with respect to freedom. See also Allison 1990). What makes this terminological ambiguity even more confusing is that early in his thinking Kant seems to have thought that the freedom necessary for moral responsibility could be established within empirical psychology. In a lecture from the 1770s, for example, Kant claims that 'practical or psychological freedom ... is treated of in empirical psychology, and this concept was also sufficient enough for morality' (28:267). Thus at least in this early lecture, the freedom that Kant discusses within his empirical psychology is both an empirical property of human beings and sufficient for—rather than merely a marker of—moral responsibility. And although Kant gives up this argumentative strategy in later lectures, the terminological confusion remains. What is more, as Kant develops his transcendental idealism, he often uses the discussion of (empirical) freedom in his empirical psychology as a starting point for discussing his transcendental philosophy. The result of these shifting views and ambiguous terminology is that it is often difficult to distinguish the perspective from which Kant is speaking at any given time, and this makes it look as though what are really only markers of moral responsibility are necessary or sufficient criteria of it. Still, it is possible to distinguish between Kant's empirical accounts and his transcendental ones at least to the extent that a reasonable Kantian view can be reconstructed.

<sup>13</sup> Here again it is important to recognize that Kant's discussion is an empirical one. Brian Jacobs puts the point well in the context of discussion of the nature of freedom of the higher faculty of desire (the 'will') in Kant's *anthropology*: 'The "*arbitrium liberum*" that Kant posits against the animalistic "*arbitrium brutum*" ... is a practical empirical concept and one that is observable when a human being resists acting solely according to the "pathological" necessity that characterizes animal will' (Jacobs, 2003, p. 120).

<sup>14</sup> Strictly speaking, he argues that it should be referred to the 'philosophical faculty' but only because 'the question ... is purely psychological' (7:213–214). Thus it is clear that the philosophical faculty has jurisdiction here only insofar as it is involved in empirical psychology.

<sup>15</sup> Observers of animals may be able to discern principles of their action, but the animals themselves discern no such principles.

<sup>16</sup> The role of the higher faculty of cognition as a marker of moral responsibility also shows up in Kant's discussion of character. Character, which is a 'sign' (7:285) or 'marker' (25: 1156) of freedom, is defined as 'that property of the will by which the subject has tied himself to certain practical principles' (7:292). But 'principles' come only from the higher faculty of cognition and thus relate only to the higher faculty of desire. One who lacks a properly functioning understanding cannot formulate action-guiding principles at all, much less 'tie himself' to them. Thus the higher faculty of desire, and with it the higher cognitive faculties, become an important precondition of, and thereby marker for, character. The connection between character and the higher faculty of desire is discussed in more detail in Frierson (2005a).

<sup>17</sup> There is no reason to see the natural faculty of understanding and the capacity to be moved by pure reason as equivalent. One might have a higher faculty of desire that is immediately determined to act by various principles of the understanding, but those principles might all be ultimately tied to various non-rational desires. Hume's notion that reason is the slave of the passions represents just such a view, and Kant nowhere claims that the mere fact of intellectual impelling causes shows that purely intellectual motives can move human beings.

<sup>18</sup> This suggestion is confirmed elsewhere. In a late lecture on metaphysics (*Vigilantius*, 1794–1795), Kant introduces moral categories into a discussion of empirical psychology by pointing out that human beings always have a capacity for action from 'pure power of choice' (my emphasis) and insists that 'a representation ... of the law of duty is always concurring alongside [any action] ... because otherwise one would make a human being equal to cattle or the devil' (29:1015). Kant seems to think here that a morally relevant feature of human beings that distinguishes us from animals (and devils) is our capacity for being motivated purely intellectually.

moral law' in these contexts (see also 4:400, 5:72–89), he seems to lay out an empirical cause of morally good action. In the second *Critique*, for example, he explains that his account will show 'what happens to the human faculty of desire as an effect of [the moral law as a] determining ground' of one's actions (5:72). The elaborate account offered there of the moral law as an incentive is congruent with Kant's more general accounts of the pure intellectual faculty of desire, whereby pure practical reason can cause various actions. In both the *Groundwork* and the second *Critique*, however, Kant's discussion of respect is not primarily an account of the empirical correlates of moral responsibility. Rather, it is an account of the experience of morally good action. Thus although these texts provide important empirical discussions of the experience of acting from duty, they do not explicitly address the criteria for determining whether an entity is morally responsible. Given this emphasis on the nature of the good will, the emphasis on the higher faculty of cognition that is found in Kant's more properly empirical works (his *Anthropology* and related lectures) should be taken as Kant's basic account of empirical markers for moral responsibility. For some purposes (especially when laying out the nature of a good will), however, it can be appropriate to supplement this basic account with an emphasis on the capacity to feel moral respect.

We have, then, the beginning of an answer to the question of when a human being (or other animal) is morally responsible. Insofar as the cause of an action can be traced via one's desires to cognitions that lie in the understanding (in the broad sense), one might<sup>19</sup> be morally responsible. To this Kant sometimes adds a capacity for one's pure reason to be a cause of desires that give rise to actions, a capacity connected with one's consciousness of the moral law. Either empirical account explains why children and the insane are not morally responsible. Children—at least very young children—do not have sufficiently developed rational capacities to act on the basis of principles of the understanding or reason. And at least for certain cases of insanity, the possibility for one's desires to be shaped by one's understanding has been lost. And either account also ties moral responsibility to empirical features of human beings, features that are within the province of empirical psychology.

This account also can be extended to cover particular deeds<sup>20</sup> of otherwise morally responsible agents. One who drowns another due to dizziness is not morally responsible even if the person has an otherwise properly functioning faculty of understanding because the dizziness itself is something over which the understanding does not and could not have any causal influence, and dizziness has the effect of causing action without allowing for influence by the understanding. It disrupts the pathway from higher cognition to higher desire. Kant even classifies a whole sphere of mental weaknesses—affects—that may cause one to act or fail to act without moral responsibility by causing one to act without the influence of the higher cognitive faculties.<sup>21</sup> Kant explains, for example, the way that shock can incapacitate without any moral responsibility attaching to one's failure to act:

One sees a child fall into the water, who one could save, however, through a small aid, but one is so shocked that one thereby cannot do anything. Shock anaesthetizes someone such that one is thereby totally incapable of doing anything. (25:592)

In this case, one does not have more responsibility because the proper functioning of one's higher cognitive faculties, and thereby of one's higher faculty of desire, has been impeded. In general, then, Kant can answer the question of what warrants ascribing moral responsibility in particular cases by appealing to the empirical fact that the relevant agents have properly functioning higher cognitive faculties. And these empirical facts about human agents will all fall under the general province of empirical psychology.<sup>22</sup>

### 3. The insufficiency of empirical markers

In the last section, I argued based on Kant's empirical psychology that the presence of higher cognitive faculties is an important criterion for moral responsibility. My justification of this empirical marker for moral responsibility, however, might seem to have proven too much. On the account that I have offered, agents are morally responsible if they are motivated by various higher cognitions and can be motivated by the moral law. I have insisted that this is merely a *marker* for moral responsibility; it is neither a necessary nor sufficient condition. But one might question whether an agent could ever be morally responsible if they *cannot* be motivated by the moral law, or if their actions are truly instinctual, caused by immediate sensations with no endorsement by higher faculties of cognition and desire (as in the case of shock discussed in the last section). Similarly, one might think that someone who meets these criteria *must* be responsible. That is, one might think that these 'markers' must be necessary and sufficient conditions of moral responsibility. In this section, I take up the question of whether these markers are sufficient for moral responsibility. In the next, I address the issue of necessity.

In his lectures on ethics, Kant is explicit that higher faculties are not sufficient for moral responsibility. There Kant says,

The ground of the fact that man is an accountable being lies

- a. not simply and solely in the fact that he is a rational being; accountability will, indeed, be founded *a posteriori* on that, but *a priori* is can still be separated therefrom. The idea is acceptable *a priori* that man, by virtue of his rational capacity, can reflect upon the grounds and consequences of his action, without his morality having to be connected with that ...
- b. absolutely necessary in addition, that he act with freedom, indeed it is only when considered as a free being that he can be accountable. (27:559, *Ethics Vigilantius*)

As Kant makes clear here, when it comes to a posteriori ascriptions of moral responsibility, one can turn to the fact that human beings are rational. This reflects the fact that rationality is a legitimate empirical marker of moral responsibility. But the empirical fact that

<sup>19</sup> As already noted, Kant is not wholly consistent about whether this capacity to be motivated by higher cognitions in general is a sufficient marker of moral responsibility or whether one must also have a capacity to be motivated by the moral law. In practice, where one is seeking to discriminate between cases, he seems to see the two criteria as equivalent.

<sup>20</sup> I avoid the term 'actions' because some have construed this to imply moral responsibility analytically.

<sup>21</sup> For more on the nature of affects, see Sorenson (2002); Frierson (2003); Borges (2004).

<sup>22</sup> One might add to this a requirement that one have the capacity for actions to be motivated by the pure higher cognitive faculties, but generally for Kant the work of distinguishing cases rests merely on the presence of a higher faculty of desire, and that in turn on the presence of a functioning understanding. The capacity for action to be caused by pure reason does not play a prominent role in the practical task of explaining who can be excused from moral responsibility. Thus the ambiguity in Kant's own discussions of determination by pure reason is not particularly important in the concrete application of empirical tests for moral responsibility. The emphasis on the understanding more generally also helps deal with Kant's reservations about the possibility of empirically determining whether an action is the result of pure reason. Admittedly, Kant insists that empirical evidence for the psychological features that correlate with moral responsibility will never be decisive, both because of the difficulty of investigating human beings (see also 7:121, 131–134, 25:1212) and because in principle empirical investigation can never yield apodictic certainty (see 4:471, B3, B124/A91). But the limitation on knowledge of freedom in the first *Critique* is more stringent than these limitations on empirical psychology.

one is rational is *insufficient* to justify moral responsibility philosophically, because one could be rational without being accountable, *if one lacks freedom* (which here refers to transcendental freedom).<sup>23</sup> And precisely because these empirical markers are empirical, they cannot provide any direct evidence that a person has that *transcendental* freedom that is the necessary condition of the possibility of moral responsibility.

The basic argument against considering these empirical markers to be sufficient conditions of moral responsibility has three steps. The first step is to show that the markers of moral responsibility that I have discussed are empirical characteristics that fit into a series of natural causes and effects.<sup>24</sup> Kant makes the empirical nature of these markers clear throughout his writings. In the *Anthropology*, the 'understanding' that indicates culpability is a 'natural faculty' (7:213). And in the first *Critique*, Kant is quite explicit about the general point that the 'causality of reason in the determination of the will' that he associates with empirical freedom is 'one of the natural causes' (A803/B831). He adds that 'even though it is reason, it must nevertheless exhibit an empirical character' (A549/B577) and raises the possibility that 'reason is itself determined by further influences' (A803/B831). And in a 1793 lecture on ethics (*Vigilantius*), Kant insists,

Even one's reason, as subjected to the laws of nature, can be considered devoid of all freedom ... Man is not set free from the mechanism of nature by the fact that in his action he employs an *actus* of reason. Every act of thought or reflection is itself an occurrence in nature ... So the fact that a man is determined to action on grounds of reason and understanding does not yet release him from all mechanism of nature. (27:502–504)

<sup>23</sup> Difficulties with terminology make interpreting Kant here a bit tricky. Kant often uses terms like freedom ambiguously, such that it is not clear whether he refers to transcendental or empirical freedom. If Kant means by 'freedom' in this passage only empirical freedom, then he could simply be saying that higher *cognitive* faculties are insufficient; one must have a higher faculty of desire (hence freedom) in addition. In the context, however, the relevant 'freedom' seems to refer to transcendental freedom.

<sup>24</sup> My account of this is necessarily somewhat brief here. The claim that higher faculties of cognitions and desire fit into a series of natural causes is defended in more detail in Frierson (2005a).

<sup>25</sup> In the *Groundwork*, he points out that 'worldly wisdom' involves 'the skill of someone in influencing others so as to use them for his own purposes' (4:416 n.). And throughout Kant's anthropology—both the published *Anthropology* and his many lectures—Kant emphasizes the importance of anthropological knowledge for influencing the choices of others. Kant bemoans the fact that 'morals and preaching that are full of admonitions ... have little effect' and attributes this failure to 'the lack of knowledge of man' (25:471–472, see also 27:358). In later lectures, Kant again emphasizes the topic of influence on others:

We must trouble ourselves to form the way of thinking ... of those people with whom we have to do ... So we are taught anthropology, which shows us how we can use people to our ends. (25:1436)

In the published *Anthropology* the same theme emerges. Kant says in his discussion of the different characters of various national groups,

We are interested only in [what] would permit judgment about what each has to know about the other, and how each could use the other to its own advantage. (7:312)

The emphasis in these passages is on the ability to influence the choices and behavior of others.

<sup>26</sup> As he says there,

This happens when the other ... confronts the subject with ... the moral law by which he ought to act. If this confrontation makes an impression on the agent, he determines his will by an Idea of reason, creates through his reason that conception of his duty which already lay previously within him, and is only quickened by the other, and determines himself accordingly to the moral law. (27:521)

In his second *Critique*, Kant goes further and offers details about how one might, through depicting the moral law in a particular way, prompt another person to act from an appreciation for the value of morality. He explains how one can

show in an example the mark by which pure virtue is tested and, representing it as set before, say, a ten-year-old boy for his appraisal, see whether he must necessarily judge so of himself ... One tells him the story of an honest man whom someone wants to induce to join the calumniators of an innocent but otherwise powerless person ... Then my young listener will be raised step by step from mere approval to admiration, from that to amazement, and finally to the greatest veneration *and a lively wish that he himself could be such a man* (though certainly not in such circumstances) ... All the admiration *and even the endeavour to resemble this character*, here rest wholly on the purity of the moral principle ... Thus morality must have more power over the human heart the more purely it is represented. (5:155–157; my emphasis)

This passage shows how the higher faculty of desire can be shaped by education in very concrete ways, and confirms that even these higher faculties are part of an empirical series of causes and effects.

Kant emphasizes the empirical nature of this potential for motivation by moral reason in the second *Critique*. He introduces his account of educating the ten-year-old boy by claiming that 'We will ... show, *by observations* anyone can make ... this property of our mind, this receptivity to a pure moral interest and hence the moving force of the pure representation of virtue' (5:152; my emphasis). Just as one presents empirical reasons for believing in various human instincts, propensities, and inclinations, so one can show *by observation* that human beings have an innate possibility for moral motivation. And the receptivity that one finds in human beings for such motivation is precisely a receptivity to *empirical influences*, such as the telling of a vivid story of moral virtue. And this receptivity is itself the *result* of empirical causes, though human science may never be able to discern precisely how it arose from them.

<sup>27</sup> Kant even suggests, in the *Anthropology*, that the higher mental faculties that correlate with moral responsibility could be reduced, in principle at least, to purely biological explanations. When he argues in favor of making decisions about sanity on psychological rather than medical grounds, his reason is that 'physicians and physiologists in general have not reached a deep enough understanding of the mechanical element in man so that they could explain, in terms of it, the seizure that led to the atrocity, or foresee it (without dissecting the body)' (7:214). Here at least, Kant leaves open the possibility that such a 'deep enough understanding' might at some point be possible.

<sup>28</sup> For details about the role of politeness in moral cultivation, see Brender (1997, 1998); Frierson (2003, 2005b).

<sup>29</sup> For discussion of the way in which the beautiful cultivates this love, see Guyer (1993); Allison (2001); Loudon (2000).

<sup>30</sup> For more on the role of unsocial sociability, see Anderson-Gold (2001); Wood (1991, 1999).



stability provides a context within which 'the citizens' inclination to violence against one another is powerfully counteracted by a greater force, namely that of government ... [by which] the development of the moral predisposition to immediate respect for right is actually greatly facilitated' (8:375 n.).<sup>31</sup> In all of these cases, social and cultural influences give rise to various patterns of intellectual desire, and these patterns can in turn explain individual human choices.<sup>32</sup> For Kant, the higher faculties of cognition and desire are not free from determination by empirical influences.

Given that the empirical markers of moral responsibility—the faculties of understanding and the higher faculty of desire—can be explained naturalistically, there are two further steps to show that these markers cannot be sufficient conditions for moral responsibility. The next step is to argue that anything that fits into a natural series of causes and effects cannot be equivalent to transcendental freedom. This is quite straightforward, because Kant defines transcendental freedom as 'independence from everything empirical and so from nature generally' (5:97). Finally, one must argue that transcendental freedom is a necessary condition of moral responsibility. Although this is complicated by the fact (noted above) that Kant changes his mind about the conditions of possibility of moral responsibility, by the time of the second *Critique* Kant makes quite clear that only transcendental freedom is sufficient for moral responsibility.<sup>33</sup> As he says there, 'without [transcendental] freedom ... no moral law is possible and no imputation in accordance with it' (ibid.). Because the empirical markers of moral responsibility do not establish this transcendental freedom, they are insufficient *proof* of moral responsibility, even if, as Kant says, they are sufficient *a posteriori* criteria.

Even with this abstract account of why empirical markers are not sufficient for moral responsibility, it might just seem implausible that one who is empirically capable of being motivated by the moral law is not morally responsible. Fortunately, Kant considers at least two hypothetical cases within which human beings would not be morally responsible despite a psychological account of action that includes higher faculties of cognition and desire. First, in the first *Critique*, Kant considers the possibility that transcendental idealism is false, that there is an incompatibility between nature and freedom. He first reiterates 'that morality necessarily presupposes freedom' but then raises the possibility—contrary to fact—that speculative reason had proved that freedom cannot be thought at all' (Bxxix). In other words, speculative reason might have shown, and many people both in Kant's day and our own think it has shown, that given natural necessity, freedom is simply unthinkable. Kant develops his transcendental philosophy in part to show that this inference from causal necessity to the impossibility of freedom is unfounded, but he here entertains the possibility that his arguments for transcendental idealism fail. And in such a case, he claims, 'freedom and with it morality ... would then have to give way to the mechanism of nature' (Bxxix). As he says later in the first *Critique*, 'If appearances are things in themselves, then freedom [and by implication morality] cannot be saved' (A536/B564). Thus even if human beings have the psychological capacity to be empirically determined by higher faculties of cognition and

desire, and even if human beings can sometimes be empirically determined by *pure rational cognitions*, unless this empirical determination is itself grounded in transcendental freedom, human beings are not morally responsible. That is, an argument from empirical psychology, though valuable for picking out particular cases, is dependent upon and secondary to Kant's successful critique of pure reason.

In the *Religion*, Kant raises a second scenario within which human actions might empirically act from higher faculties but within which humans would not be morally responsible. There Kant addresses the problem that human beings are 'radically evil', corrupt at the level of our most fundamental maxims. He suggests that 'some supernatural cooperation is ... needed for his becoming good or better' (6:44), but in order to save moral responsibility, Kant must carefully restrict the scope of this intervention. He argues,

The concept of a supernatural intervention into our moral though deficient faculty ...—this is a transcendent concept, merely an idea of whose reality no experience can assure us.—But even to accept it as idea for purely practical intent is very risky and hard to reconcile with reason; for what is to be accredited to us as morally good conduct must take place not through foreign influence but only through the use of our own powers. (6:191)

Here Kant entertains the possibility that God might bring about a moral shift in one's fundamental maxims through an act of grace, but he raises the practical problem that insofar as God—a 'foreign influence'—causes this shift, it cannot be accounted to oneself and thus cannot constitute true *moral* goodness. As a result, Kant insists that 'the human being must ... make himself antecedently worthy of receiving' grace (6:44). For the purposes of this paper, I am not interested in the specific dynamics of Kant's account of grace. What is important here is only that if God caused a change in one's fundamental moral maxims, the effect of this change would be that one would more consistently act from principles of pure reason. One would satisfy the empirical criteria Kant sets out for moral responsibility. But because the ultimate ground of one's actions would lie in God, rather than in one's own transcendental freedom, one would not *in fact* be morally responsible, in that one's morally good conduct could not be accredited to one.<sup>34</sup> As in the previous case, there are no indicators within empirical psychology to distinguish a world of intrusive grace that precludes freedom from a world of supplementary grace that supports it.

Both of these hypothetical cases have in common that Kant's empirical account of human agency is left unchanged, but that account is shown to be ultimately grounded not in human freedom but in something else—natural laws or God's grace.<sup>35</sup> In these cases, despite the fact that human beings can act on the basis of rational principles, and even on the basis of pure principles of practical reason, such action is not free because it is not *noumenally* free. And this lack of transcendental freedom is sufficient for Kant to deny moral responsibility without any change in his empirical psychology. Thus

<sup>31</sup> Most of these accounts focus on the gradual development of the natural predispositions, and in particular *moral* predispositions, in the human *species*. For more on the nature of this development of moral predispositions, see Frierson (2003), pp. 152–162; Munzel (1999); Anderson-Gold (2001); Wood (1991, 1999).

<sup>32</sup> It is important not to lose sight of the limits of these empirical accounts. Insofar as the higher faculty of desire is studied within empirical psychology, its choices are to be explained like any other natural event—by reference to natural causes. But human choice cannot *ultimately* be explained by this natural story. Human choices are also transcendently free, and in *that* sense, they are the grounds rather than the effects of any empirical story that might be told. Moreover, any human choice can be thought of from a practical perspective, from which perspective such a choice cannot be considered a mere result of empirical causes. And when one makes a choice on the basis of pure practical reason, such a choice is specifically free from empirical determination—but free only from within the practical perspective.

<sup>33</sup> Needless to say, the fact that Kant claims that transcendental freedom is necessary for moral responsibility is not enough to establish that this claim is correct. For the purpose of articulating an overall picture of Kant's account of empirical markers, I take this claim for granted here.

<sup>34</sup> Of course, one could still attribute a person's moral wickedness to him because God would not be the ultimate cause of that wickedness. In that sense, Kant need not remove moral responsibility for the radical evil that makes grace necessary, but he does remove moral responsibility (credit) for the revolution that shifts one from evil to good.

<sup>35</sup> These two cases might not be that different, since for Kant God determines the natural laws, at least to a considerable degree. (see 5:124–131).

the empirical psychology that serves as a marker for moral responsibility is not *sufficient* for moral responsibility.<sup>36</sup>

#### 4. The contingency of empirical markers

Even if no empirical features of human psychology are *sufficient* for showing that one is morally responsible, one might think that at least some empirical features are *necessary* for moral responsibility. That is, those who act on the basis of principles and have the capacity to act on pure principles of reason might not be morally responsible if their behavior is ultimately determined by God or natural laws, but there is no way to hold morally responsible a person who cannot act on the basis of principles at all.

As far as I know, Kant never explicitly discusses the possibility of an entity that is morally responsible but lacks an empirical psychology that includes a higher faculty of desire and the capacity for this desire to be determined by pure principles of practical reason. (Even God, it seems, would meet these two criteria.) Nonetheless, Kant's transcendental idealism and his account of transcendental freedom open the possibility for ascribing moral responsibility to entities regardless of their empirical psychology.

Kant's argument for transcendental freedom in human beings shifts between the *Groundwork* and the second *Critique*, so here I will focus on the structure of the argument in the second *Critique*. That argument takes place in the context of the transcendental idealism that Kant defends in the first *Critique*. In particular, the Third Antinomy of the first *Critique* raises the possibility that 'there is no freedom' because 'everything in the world happens solely in accordance with the laws of nature' (A445/B473). Kant responds to this possibility by claiming that although 'every effect in the world must arise ... from nature', yet there is still the possibility that 'both [nature and freedom], each in a different relation, might be able to take place simultaneously in one and the same occurrence' (A536/B564). Kant develops his account of this possibility in terms of the notion of 'character'. As he says,

Every effective cause must have a *character*, i.e., a law of its causality, without which it would not be a cause at all. And then for a subject of the world of sense, we would have first an *empirical character*, through which its actions, as appearances, would stand through and through in connection with other appearances in accordance with constant natural laws, from which,

as their conditions, they could be derived ... Yet second, one would also have to allow this subject an *intelligible character*, through which it is indeed the cause of those conditions as appearances, but which does not stand under any conditions of sensibility and is not itself appearance ... In its empirical character, this subject, as appearance, would thus be subject to causal connection ... But in its intelligible character ... this subject would nevertheless have to be declared free ... (A539/B567)

Kant does not think that his distinction between empirical and intelligible character is sufficient to *prove* that the 'subject' is in fact free. He does not here establish 'the reality of freedom ... [but only] that nature at least *does not conflict with ... freedom*' (A558/B586). But what is crucial for my purposes is that the account that establishes that nature does not conflict with freedom does not depend on any particular account of the sorts of natural causes at play.<sup>37</sup> Kant's point is simply that natural necessity of any kind can be grounded in an intelligible character that can be transcendently free.<sup>38</sup>

In that context, there is nothing *preventing* the ascription of an intelligible character, and with it transcendental freedom, to any empirical objects at all. To use Hume's famous example, it is possible—in a very weak sense—that the 'sapling ... which ... at last overtops and destroys the parent tree' (Hume, 1978, p. 467) is transcendently free (and thus potentially both morally responsible and guilty of parricide!). The invocation of 'character' even highlights this possibility because Kant is explicit that 'every effective cause must have a character' (A539/B567), including the causes governing the development of trees as much as the causes governing human behavior.<sup>39</sup> Of course, the arguments in the first *Critique* merely show the *compatibility* of freedom with natural laws. They *establish* freedom for neither saplings nor human beings. But the claim of compatibility works equally well for both cases.

Kant's positive argument for human freedom comes in the second *Critique*.<sup>40</sup> In perhaps the most intuitive statement of the structure of his argument, Kant explains that one 'judges, therefore, that he can do something because he is aware that he ought to do it' (5:30). This argument begins from the fact that one 'ought to do' things. From that 'fact' of moral obligation—the moral law itself—one judges that one is free because freedom is a condition of the possibility of moral responsibility.<sup>41</sup> Because the moral law applies to

<sup>36</sup> This also means that a skeptic who claims that human beings are in fact causally determined all the way down and thus not morally responsible cannot be refuted by Kant's defense of human freedom and moral responsibility. As long as the skeptic objects on grounds like the ones mentioned here, nothing in either Kant's transcendental idealism or in his empirical psychology can *prove* that human beings are morally responsible. For more on the 'modesty' of Kant's system, see Ameriks (2000), pp. 1–77.

<sup>37</sup> Admittedly, there are some passages in the first *Critique* that suggest that Kant ties transcendental freedom to a particular human faculty: reason. A full discussion of these texts is beyond the scope of the present paper. Many of these are compatible with the account developed here, but others reflect Kant's early argument (developed in detail in the *Groundwork*) for freedom, and argument that he rejects in the second *Critique* (see Ameriks 2003).

<sup>38</sup> Some of the same conclusions about the possibility for moral responsibility for non-rational entities can be drawn from Kant's account of maxims. A maxim, according to Kant, is a 'subjective principle of volition' (4:400 n.), a principle on the basis of which one chooses to act. This notion of a maxim, however, is not primarily a concept of empirical psychology, but of practical philosophy. Kant primarily discusses the concept of a maxim as part of general discussions of moral evaluation and deliberation, and only very rarely discusses maxims in the context of his empirical psychology. A maxim is the principle that provides the actual basis of one's action, whether or not one can discern such a principle by empirical means, including introspection. Given that one cannot experience this ultimate basis of action, however, there is no necessary reason to preclude non-humans from acting on maxims. Of course, it is reasonable to distinguish other entities from humans on the grounds that only humans have a first-person perspective within which talk of maxims makes sense, but even this claim is based on assumptions about what is necessary for a first-person perspective, assumptions that Kant's transcendental idealism helps undermine.

<sup>39</sup> Kant does distinguish between this character in a general sense and character in the narrower sense appropriate to human moral responsibility. But the account in the first *Critique* leaves open the possibility of ascribing moral responsibility to any entity with character in the broad sense, even if Kant himself limits responsibility to those with his narrower sense of 'character simply' (*Character schlechthin*, see also 7:285).

<sup>40</sup> Kant gives some arguments for freedom in the first *Critique* and in the *Groundwork*, but for the purposes of this paper, I am taking the second *Critique* to offer Kant's most mature argument for transcendental freedom. The second *Critique* does depend on the first, but only to establish the compatibility of freedom with natural necessity, not for any positive arguments for freedom. The precise details of Kant's argument in the second *Critique* are controversial, and I will not defend my particular reading of it here. Broadly speaking, I follow Karl Ameriks in seeing it as a regressive argument, based on the premise that 'the moral law ... is simply given' (Ameriks, 2003, p. 53).

<sup>41</sup> This approach contrasts with Allison, who argues that 'the fact is best construed as the consciousness of standing under the moral law and the recognition of this law' (Allison, 1990, p. 233). I take this fact to be too psychological. When Kant describes the fact as, for instance, 'consciousness of the moral law' (5:31), I take him to refer to the moral law itself as the fact; of which one is conscious. Similarly, one might say, 'from the awareness that the sky was getting darker, she concluded that night was approaching'. But no one thinks that one reasoned from 'the awareness' to the approach of night. Rather, the relevant premise is 'the sky is getting darker'. Likewise when Kant refers to 'consciousness of the moral law' as the fact of reason, he is elliptically referring to the moral law itself, or better the fact that there are moral laws for us, that we are obligated to act in certain ways rather than others.



human agents, humans must be morally responsible. Therefore, humans must be transcendently free.

Now if we assume that this argument works in the case of human beings, we are not committed to saying that it will work in the context of the parricide sapling. In that sense, Kant is *not* committed to the apparently absurd conclusion that saplings are free and thus blameworthy. Kant makes clear, in fact, that his own view is that 'from all our experience we know of no being other than a human being that would be capable of obligation' (6:442).<sup>42</sup> But the reason that the argument does not apply to the sapling raises some concerns about Kant's overall argument. Because saplings fail to exhibit any of Kant's empirical markers for moral responsibility, saplings are not morally responsible.<sup>43</sup> And because the basis of ascribing transcendental freedom to human beings is that they are morally responsible, there is no basis for ascribing it to trees. But now it should be clear that this argument poses a problem for the claim that empirical markers can be shown to be *necessary* for moral responsibility. Kant can show that the sapling is not transcendently free only by *assuming* that his empirical markers are reliable criteria for moral responsibility. If anyone challenges this assumption, none of Kant's arguments in the first or second *Critique* will be sufficient to prove that one cannot hold saplings responsible, and the overall structure of those *Critiques* will even provide a systematic account of how to make sense of such moral responsibility.<sup>44</sup> Thus any argument to show that Kant's empirical markers are necessary for moral responsibility will be circular.

## 5. Common sense conclusion

So what, then, are the status of these empirical markers, if they are neither necessary nor sufficient conditions of moral responsibility? And what role does empirical psychology actually play in determining the scope of moral responsibility? The answer is that empirical markers are just that, markers, and as markers they are both necessary and sufficient. That is, for any particular case where one questions the moral status of a particular entity or an entity acting in a particular context, one need only discern whether or not the entity has active and functioning higher cognitive faculties. If the entity does, then it is morally responsible; if not, then it is not responsible. And because the detection of these cognitive faculties is an empirical question, empirical psychologists will settle the question of moral responsibility in any particular case. The denial that these empirical conditions are necessary or sufficient is only a denial that they are *strictly* necessary or sufficient. That is, the world might have been set up in such a way that these empirical features are not markers for moral responsibility. But because the world is not set up in that different way, these markers are, for all practical purposes in particular cases, necessary and sufficient.

But how does Kant know that the world is not set up so that higher cognitive faculties are merely an illusion, or so that all entities—from stones and trees to humans and angels—are morally

responsible? As I have argued, there is no basis in Kant's transcendental idealism for picking out higher cognitive faculties, nor can there be any argument within empirical psychology. But Kant does not claim that transcendental idealism or empirical psychology alone provide a sufficient basis for *any* moral claims. So it should not be surprising that they do not provide a sufficient basis for this one. Instead, Kant turns to 'common rational cognition' (4:393) for a starting point to moral enquiry. In an important footnote at the beginning of the *Critique of practical reason*, Kant responds to a critic who accuses him of having nothing new to say about ethics. Kant asks, 'Who would even want to introduce a new principle of all morality and, as it were, first invent it? Just as if, before him, the world had been ignorant of what duty is or in thoroughgoing error about it' (5:9 n.). Kant sees his moral philosophy, and arguably his transcendental philosophy as a whole, as providing a rigorous philosophical defense of moral claims that 'the world' has long known.<sup>45</sup> This humility about his own project can be ascribed, at least in part, to Kant's reading of Rousseau. Kant describes his encounter with Rousseau in a now famous journal entry:

There was a time when I ... despised the rabble who knows nothing. Rousseau set me right. This blind prejudice vanishes; I learn to respect human nature. (20:44)

As I have described Kant's argument for human freedom, it is an account that defends and justifies what Kant takes to be common human convictions about the nature of moral responsibility and obligation. And Kant's account of empirical markers for moral responsibility is, similarly, an attempt to provide a systematic philosophical account that captures what 'the rabble' already thinks about moral responsibility. Thus Kant would defend his claim that one cannot be held responsible for (most) emotions by appealing to common sense.

This suggests an important Kantian insight into the proper relationship between the 'common understanding' and those sciences—psychology in particular—most closely tied to human life. Immediately preceding the previous quotation in which Kant describes his relationship to Rousseau, Kant writes that 'One must teach youth to honor the common understanding as much for moral as for logical reasons' (ibid.). This paper has suggested some of the logical reasons why one should honor the common understanding. Empirical psychology, even with the help of transcendental philosophy, cannot determine which empirical features are proper markers of moral responsibility. This determination must come from common sense moral judgment and practice, from our everyday experiences of moral responsibility in deliberation and moral evaluation. But transcendental philosophy can provide a framework for integrating those common sense moral judgments with empirical psychology, and in that context, psychology can help us in difficult circumstances to determine whether the relevant features are present. And this, 'as much for moral as for logical reasons', is all that one should require.<sup>46</sup>

This problem is not devastating for Kant's philosophy, nor even for his account of empirical markers for moral responsibility. Argu-

<sup>42</sup> Elsewhere Kant speculates about the possibility of 'rational beings' 'on some other planet' (7:332) and at times suggests that he believes that there are such beings (see e.g. 1:365–366). For the purposes of this paper, Kant's speculations about extraterrestrial life are inessential. There is no reason to believe that Kant would use different criteria for moral responsibility for extraterrestrial life than for terrestrial life.

<sup>43</sup> There is another less consequential reason that the argument of the second *Critique* does not quite apply to saplings. Kant's argument is *addressed* to an intelligent practical reasoner. And of course to 'judge that one is free', one must have higher cognitive faculties, because without those, one can judge nothing at all. What this point establishes, however, is only that no sapling can believe that it is free, since no sapling has beliefs. It shows nothing about whether or not a sapling is in fact free.

<sup>44</sup> It will, of course, be more difficult to write an account of 'the incentives of pure practical reason' for a tree. But this shows only that the experience of being a tree (if there even is such an experience) is something humans cannot understand. It does not show that the 'character' of the tree, the fundamental principles that govern its 'activity' in the world, cannot be grounded (dare I say 'rooted') in freedom and even in the moral law.

<sup>45</sup> Karl Ameriks has helpfully described this approach in Kant as 'Kant's Modest System' (Ameriks, 2000).

<sup>46</sup> There is, admittedly, a challenge that arises in the context of competing 'common sense' moral judgments of moral responsibility, such as whether or not one can be responsible for emotions and whether or not certain animals are morally responsible. Sometimes these apparent disagreements are ultimately disagreements at the level of empirical psychology—one person thinks that certain emotions reflect higher cognitive faculties, another does not—but insofar as they are at a deeper level—about whether higher cognitive faculties are necessary at all—they suggest the need for something that goes beyond the Kantian account I have offered here.

ably, the compatibility of Kant's overall idealism with different accounts of the empirical criteria of moral responsibility is a strength of that account. Kant's philosophy does not rise or fall with any particular empirical psychology. Moreover, the flexibility of Kant's overall transcendental philosophy provides important opportunities for contemporary neo-Kantians to develop authentically Kantian theories of moral responsibility (and even of transcendental freedom) that are not identical to Kant's account. Paul Guyer provides an excellent beginning to such a theory, explaining how one could, consistent with Kant's overall philosophy, defend the claim that people are directly responsible for their emotions. One might develop similar theories to explain how one can be morally responsible for the way one perceives the world or to explain how certain kinds of animals can be held morally responsible.

But with this opportunity comes a special challenge. Kant's philosophy needs to be supplemented with an account of how one can arbitrate between competing common sense views of moral responsibility. Given several different possible Kantian theories of the empirical criteria of moral responsibility, one needs a way to distinguish between them. And Kant's idealism has cut off what might seem the most obvious way. One might claim that any action that can be explained in terms of natural necessity cannot be a matter of choice and therefore cannot be an action for which one is morally responsible. But because Kant claims that one can be free despite being 'determined' in accordance with natural laws, he cannot offer this easy answer. Instead, he depends on reaching a kind of reflective equilibrium between his theory of empirical markers and common sense views about moral responsibility. The current challenge is to develop a way to arbitrate between theories when there are conflicting trends within 'common sense views'.

## References

- Allison, H. (1990). *Kant's theory of freedom*. New York: Cambridge University Press.  
 Allison, H. (2001). *Kant's theory of taste*. New York: Cambridge University Press.  
 Ameriks, K. (1981). Kant's deduction of freedom and morality. *Journal of the History of Philosophy*, 19, 53–79.  
 Ameriks, K. (2000). *Kant and the fate of autonomy: Problems in the appropriation of the critical philosophy*. New York: Cambridge University Press.  
 Ameriks, K. (2003). Kant's transcendental deduction as a regressive argument. *Interpreting Kant's Critiques* (pp. 50–66). New York: Oxford University Press.  
 Anderson-Gold, S. (2001). *Unnecessary evil*. Albany: SUNY Press.

- Beck, L. W. (1960). *Commentary on Kant's Critique of practical reason*. Chicago: University of Chicago Press.  
 Beck, L. W. (1987). Five concepts of freedom in Kant. In T. J. Szrednicki (Ed.), *Philosophical analysis and reconstruction: Festschrift for Stephen Körner* (pp. 35–51). The Hague: M. Nijhoff.  
 Borges, M. (2004). What can Kant teach us about emotions. *Journal of Philosophy*, 103, 140–158.  
 Brender, N. (1997). *Precarious positions: Aspects of Kantian moral agency*. Ph.D. thesis, Johns Hopkins University.  
 Brender, N. (1998). Keeping up appearances: Kantian moral agency and the fragility of respect. Paper presented at the NAKS Session of the APA Pacific Division.  
 Frierson, P. (2003). *Freedom and anthropology in Kant's moral philosophy*. New York: Cambridge University Press.  
 Frierson, P. (2005a). Kant's empirical account of human action. *Philosophers' Imprint*, 5(7), 1–34.  
 Frierson, P. (2005b). The moral importance of politeness in Kant's anthropology. *Kantian Review*, 9, 105–127.  
 Guyer, P. (1993). *Kant and the experience of freedom: Essays in aesthetics and morality*. New York: Cambridge University Press.  
 Hatfield, G. (1992). Empirical, rational, and transcendental psychology. In P. Guyer (Ed.), *The Cambridge companion to Kant* (pp. 200–227). New York: Cambridge University Press.  
 Hume, D. (1978). *A treatise of human nature* (2nd ed.) (L. A. Selby-Bigge, Ed.; rev. by P. H. Nidditch). Oxford: Oxford University Press. (First published 1739)  
 Jacobs, B. (2003). Kantian Character and the problem of a science of humanity. In B. Jacobs & P. Kain (Eds.), *Essays on Kant's anthropology* (pp. 105–134). New York: Cambridge University Press.  
 Kain, P. (n.d.). Kant's conception of human moral status. Unpublished manuscript.  
 Kant, I. (1900–). *Kants gesammelte Schriften* (Königlich Preussischen Akademie der Wissenschaften, Ed.) (29 vols. to date). Berlin: G. Reimer [later W. de Gruyter].  
 Kant, I. (1995–). *The Cambridge edition of the works of Immanuel Kant in translation* (H. Allison, R. Brandt, P. Guyer, R. Meerbote, C. D. Parsons, H. Robinson, J. B. Schneewind, & A. W. Wood, Eds.) (17 vols. to date). Cambridge: Cambridge University Press.  
 Kuehn, M. (2001). *Kant: A biography*. New York: Cambridge University Press.  
 Loudon, R. (2000). *Kant's impure ethics: From rational beings to human beings*. New York: Oxford University Press.  
 Munzel, G. F. (1999). *Kant's conception of moral character: The 'critical' link of morality, anthropology, and reflective judgment*. Chicago: University of Chicago Press.  
 Sloan, P. (2002). Preforming the categories: Eighteenth-century generation theory and the biological roots of Kant's *a priori*. *Journal of the History of Philosophy*, 40, 229–253.  
 Sorenson, K. (2002). Kant's taxonomy of the emotions. *Kantian Review*, 6, 109–128.  
 Sturm, T. (2001). Kant on empirical psychology: How not to investigate the human mind. In Watkins (Ed.), *Kant and the sciences* (pp. 163–184). Oxford: Oxford University Press.  
 Wood, A. (1991). Unsociable sociability: The anthropological basis of Kantian ethics. *Philosophical Topics*, 19, 325–351.  
 Wood, A. (1999). *Kant's ethical thought*. New York: Cambridge University Press.  
 Zammito, J. (2003). This inscrutable principle of an original organization: Epigenesis and 'looseness of fit' in Kant's philosophy of science. *Studies in History and Philosophy of Science*, 34, 73–109.