In: *Kant on Emotion and Value*, ed. Alix Cohen.
Basingstoke: Macmillan, 2014.
pp. 146-165.

# 7

# Debunking Confabulation: Emotions and the Significance of Empirical Psychology for Kantian Ethics

*Pauline Kleingeld*

It is frequently argued that research findings in empirical moral psychology spell trouble for Kantian ethics. Results from psychology and neuroscience, in particular, have been used to argue that human moral judgment and behavior are pervasively influenced by emotional triggers and inhibitors. Some argue for the strong thesis that our behavior is '*typically*' determined by emotional responses to situational factors, rather than by rational processes, and that even our cognitive processes are best explained in terms of emotional responses to features of the situation, rather than in terms of rational deliberation.[1] Others argue for a more restricted claim, namely, that the empirical research does not show reason to be ineffective in general, but rather that it debunks Kantianism in particular. Sometimes the charge is merely that Kantianism is mistaken about how human beings work, but it has also been argued that Kantianism should itself be understood as the product of precisely the emotion-driven processes it fails to acknowledge. The charge, then, is that Kantian moral theory as such is best understood as the result of emotional gut reactions. This claim has been formulated most prominently by Joshua Greene, who argues that despite Kantianism's rationalist ambitions, *emotion* underlies not only the deontological judgments of ordinary people but also the theoretical justifications of deontology by Kantian moral philosophers.[2]

In an article entitled 'The Secret Joke of Kant's Soul', Greene writes, for example, that empirical psychology 'casts doubt on deontology as





a school of normative moral thought' (Greene (2008), 67). He writes further that empirical research suggests that

> what deontological moral theory really is, what it is *essentially*, is an attempt to produce rational justifications for emotionally driven moral judgments, and not an attempt to reach moral conclusions on the basis of moral reasoning. (Greene (2008), 39)

'Strictly empirical' claims regarding psychological processes show that Kantianism is 'an exercise in moral rationalization' and 'a kind of moral confabulation' (Greene (2008), 36, 63). On Greene's view, ordinary deontological judgments can be traced back to emotional reactions; and deontological theory results from the general human tendency to strive to find rational explanations for everything, and to make them up when none can be found. What Greene argues is not so much that there are logical flaws with specific Kantian arguments as that Kantian ethics is (and is '*essentially*') a form of rationalization and confabulation. As Greene puts it: 'you can spot a rationalizer without picking apart the rationalizer's reasoning' (Greene (2008), 67).

Kantian moral theorists seem relatively unconcerned about such charges. This is not only because the characterizations of Kantianism are replete with caricatures, but, more importantly, because there seems to be a short response that is so obvious that it is hardly worth spelling out. This is the rejoinder that neither Kant nor Kantians claim that humans do act fully rationally, and that their point is not descriptive but normative. Because on the Kantian account the normative criterion for moral agency is a *rational* principle, the response continues, empirical facts about moral agency are *normatively irrelevant* at the most fundamental level. Just as mathematicians tend to regard empirical studies of mathematical problem-solving behavior as having no bearing on the validity of mathematical proofs, Kantians tend to see empirical studies of moral behavior as theoretically uninteresting to their philosophical project. Moreover, strictly speaking, not even the most compelling evidence that humans act merely on their emotions could ever prove Kantianism wrong. It is impossible with certainty to infer a person's inner disposition on the basis of their outward behavior or the neurological processes going on inside their skull. Even if we could gain such certainty, and if all we ever observed in ourselves and others was emotion-driven behavior, this still would not invalidate the normative claim that we *ought* to do the right thing for the right reason, that is, that we ought to act in accord with duty, from duty.

Yet there are three reasons for a more thorough engagement with the empirical research and its implications. For one thing, the trend in moral theory is increasingly to pay attention to empirical psychology and neuroscience, and the view that the research results pose difficulties for Kantianism is widespread. If Kantians remain silent, this could be mistaken for their being 'dumbfounded'.

Second, the short and easy response could understandably be read as merely a defensive move in a rearguard battle. Not only does it fail to address the point about the alleged origins of Kantianism itself, it also reinforces the opponents' impression – however unwarranted it may be – that Kantians comfortably insulate themselves in their own *a priori* theoretical edifice, shielding themselves from the hostile world of facts. For the short reply does nothing to address the concerns and considerations of those who formulate the challenge. I believe it is important to examine the strength of their argument as such, rather than merely to claim that the attacks are not fatal in terms of one's own theory. Otherwise, one may sound, to the opponent at least, a little too much like the overconfident Black Knight in *Monty Python and the Holy Grail*, whose arms and legs are cut off one-by-one, but who keeps protesting that, in his eyes, 'Tis but a scratch'.

Third, and even more importantly, Kantian moral theorists themselves risk overlooking something morally significant if they do not engage more thoroughly with empirical research on human emotions and agency. Empirical moral psychology has moral import, even if it has no bearing on the justification of the basic principle of morality. In fact, as I shall argue, in this regard Kantian moral theorists can and should follow Kant's own lead.

In this chapter I argue for a negative and a positive thesis. The negative thesis is that the critics' 'debunking' argument is invalid because it begs the central question. The positive thesis is that Kantians can and should wholeheartedly embrace the current interest in empirical moral psychology because the empirical facts about human psychology are morally relevant. By making the case for these two claims, I hope to steer the discussion of the philosophical implications of empirical psychology away from the currently dominant focus on its potential to 'debunk'. It is more fruitful to redirect our focus toward the positive use, for moral agency, that can be made of it.

I focus first on the structure of Greene's argument. I explain the question-begging nature of his debunking argument, showing that if Kantian ethics *can* be justified, empirical evidence does not debunk it, and that if it clearly *cannot* be justified, there is nothing to debunk



(section 1). I then examine why the question-begging argument is thought to have any plausibility at all, and I trace this back to a mistaken moral intuitionist understanding of Kantian ethics (section 2). In the final section, and with the help of Kant himself, I develop the argument for the claim that empirical psychology is of greater significance for Kantian ethics than is commonly thought (section 3).

## 1 Emotions and debunking arguments against Kantianism

Greene argues that people who reach deontological moral judgments generally do so on the basis of emotional reactions, not reasoning. He argues further that philosophers who develop deontological moral theories are best understood as merely making up a pseudo-justification for these emotion-driven judgments. How exactly does Greene argue for his position, and why is it question-begging?

Greene grounds his claims in his empirical work in neuropsychology. In a seminal article (Greene *et al.* (2001)), he and his co-authors describe experiments in which subjects were asked to solve sets of moral dilemmas while in an fMRI scanner. This allowed the authors to measure response times and to determine which brain areas were involved in the process. Greene *et al.* claim to find a surprising pattern, namely, that the deontological and consequentialist answers were associated with a marked difference in response time and a difference in brain areas involved. The fMRI data seemed to indicate that deontological judgments are formed while areas of the brain associated with emotions are active, and that consequentialist judgments are formed while areas of the brain associated with cognitive control are active. Furthermore, they claimed to find that it takes those who give consequentialist answers longer to respond, in the case of personal moral dilemmas, than it takes those who give deontological answers. Elsewhere (Greene (2009)), Greene maps these patterns onto the distinction central to dual-process theory. This is the distinction between two presumed cognitive systems: one that is quick, unconscious, and emotion-driven, and another cognitive system that is slow, conscious, and controlled. He concludes that the data indicate that deontological judgments tend to stem from quick, unconscious, emotion-driven processes, whereas consequentialist judgments tend to be the result of slow, conscious, controlled cognitive processes. Greene *et al.* use this analysis to explain, for example, why most people regard it as morally wrong to push a heavy man into the path of a runaway trolley to save five others, while most also regard it as morally right to

save five people by diverting a trolley onto a different track where it will kill one person:

> The thought of pushing someone to his death is, we propose, more emotionally salient than the thought of hitting a switch that will cause a trolley to produce similar consequences, and it is this emotional tendency that accounts for people's tendency to treat these cases differently. (Greene *et al.* (2001), 2106)

In this case, the idea of pushing a person to his death is more emotionally salient than the idea of hitting a switch to divert the trolley. Consequentialists *reason* their way past their primary emotional responses to their consequentialist answer, whereas deontologists simply condemn pushing the man, on the basis of their emotional reaction. This, Greene *et al.* take the data to suggest, is just one example of a general pattern in ordinary moral judgment.

Greene extends the scope of his argument from the claim that emotions drive the deontological judgments of test subjects to the claim that deontological *philosophy* is essentially emotion-driven. Kantian philosophers, according to Greene, are no more reason-driven than ordinary subjects who reach deontological moral judgments. For this extension of his argument, he appeals to the general human tendency to rationalize behavior, as documented in social psychology. Referring to research by Jonathan Haidt (2001) and Timothy Wilson (2002), he writes: 'Psychologists have repeatedly found that when people don't know why they're doing what they're doing, they just make up a plausible-sounding story' (Greene (2008), 61). One of the examples he mentions is that of Richard Nisbett and Timothy Wilson's classic stockings experiment. Test subjects had to select one pair of nylon stockings from a line-up of four identical pairs, without knowing that they were identical; they tended to pick the one on the right-hand side of the display, and when asked to explain their choice, they came up with various reasons for their preference, such as superior knit or elasticity. With one exception, all subjects denied any position-effect when asked about it. None of their reasons made *real* sense, however, because all samples were in fact identical (Nisbett and Wilson (1977), 243–4; Wilson (2002), 102–3). This very same phenomenon of 'confabulation' is at work in deontological philosophizing, according to Greene:

> Deontology, then, is a kind of moral confabulation. We have strong feelings that tell us in clear and [no] uncertain terms that some things *simply cannot be done* and that other things *simply must be done*. But it is not obvious how to make sense of these feelings, and so we, with the help of some especially creative philosophers, make up a rationally appealing story. (Greene (2008), 63)



In other words, deontological philosophers invent quasi-reasons to explain what is really merely an emotional reaction. Once this process is revealed, Greene assumes, the quasi-reasons should lose their grip and evaporate.

Greene *et al.*'s (2001) paper and a number of Greene's subsequent papers have sparked a wider, intense debate. Recently, it has also been subjected to fundamental criticisms, and Greene has published several replies. His research results have been called into question on conceptual and methodological grounds. Richard Dean (2010) has gone carefully through the available empirical data and shown that they are currently inadequate to undermine deontological theories. Dean argues that the evidence so far is too weak to support the claim that deontological judgments are based on emotion. Jonathan McGuire and his co-authors (2009) have pointed out that the alleged difference in response time between deontological and consequentialist answers was entirely due to a design flaw in the experiment.[3] Frances Kamm (2009) has criticized the way Greene distinguishes between 'consequentialist' and 'deontological' responses. Selim Berker has argued that Greene has not shown that neuroscience has any real 'normative significance' (Berker 2009, *cf.* Sauer 2012b). As a result of these criticisms, Greene has had to retract significant elements of his initial theory, such as his claim regarding the response time (Greene (2009)) but he still defends the debunking thesis.

Indeed, despite the methodological problems mentioned, there is no denying the fact that there is a large and growing body of empirical research detailing both the unconscious influence of emotional factors on moral judgment, and people's tendency to confabulate when they feel they have to come up with reasons (see, e.g., Prinz (2006) and Haidt (2001)). As a matter of fact, emotions *are* (at least, they often are) involved in the formation of moral judgments, and confabulation does happen; so it is still important to examine what this evidence does or could show, and what its implications are or could be in relation to Kantian ethics.

Let us, then, return to Greene's basic argument quoted above. He argues from the premise that we have strong emotional reactions ('some things simply cannot be done'; 'other things simply must be done') *via* the premises that 'it is not obvious how to make sense of these feelings'

and that 'people tend to confabulate in such cases', to the conclusion that deontological theory is nothing but 'moral confabulation'. The argument runs as follows:

P1. People have strong 'deontological' feelings regarding moral issues.

P2. People have a strong tendency to make sense of their feelings, and they confabulate when they cannot make sense of them otherwise.

P3. It is not obvious how to make sense of the strong deontological feelings.

C.: Deontology is a kind of moral confabulation.

The first and second premises are empirical, but the third is evaluative. Greene's conclusion is that deontology *is* a form of moral confabulation. This raises questions about the proper understanding of P3. Greene's statement that 'it is not obvious how to make sense' of these feelings could be taken to mean that it *takes some effort*, some further thinking, to make sense of the feelings at issue, but on that reading the conclusion clearly does not follow. After all, if 'deontological' moral theory is the right (albeit effortful) way to 'make sense' of these feelings, P3 gives us no reason to regard Kantianism as confabulation. To get to his strong conclusion, Greene must be understanding 'is not obvious' as meaning that it *is not obvious and cannot be made genuinely obvious* how to 'make sense' of deontological feelings. On this second reading of 'is not obvious', P3 would mean that there is (currently or perhaps even in principle) no convincing philosophical justification of deontological moral assumptions and the moral judgments they lead to.

On the first interpretation of 'not obvious' the conclusion does not follow; but on the second interpretation the argument becomes question-begging. For then the conclusion follows, but only because the argument already builds in the premise that Kantianism cannot be convincingly justified. Let me clarify the problem further by thinking through how Greene's debunking argument would fare in case there were a convincing argument in support of Kantian ethics.

Suppose, for the sake of argument, that a convincing argument can be given in defense of a specific Kantian moral judgment, and suppose that test subjects reach the same judgment on the basis of emotional reactions (supposing, again for the sake of argument, that this can be shown to be the case). The latter fact would not make the Kantian argument for this judgment any less convincing. It would just mean that there happens to be convergence between the conclusion of the moral



argument and the subjects' emotional reaction. Even if this convergence were regarded as an unlikely coincidence, the fact of the coincidence does not make the valid argument invalid.

Furthermore, if a convincing argument can be given, the convergence need not be regarded as a coincidence, let alone an unlikely one. If a good argument can indeed be given for a certain moral judgment, converging emotional reactions of ordinary people could be understood simply to reflect the fact that this moral insight has become socially engrained to the point of having become intuitive (Sauer (2012a)). In that case, the emotional reaction could be based on this engrained moral insight, instead of the other way around. Perhaps this relationship is an indirect one, mediated by education, tradition, and other contingent empirical factors, but this does not reduce the normative force and validity of the convincing case that can be made in defense of the judgment.

In other words, if a good argument for the Kantian position can be given, empirical evidence that test subjects reach the same conclusion in other ways (namely, through emotional reaction-patterns) does not debunk this position. On the contrary, the argument might make it possible to understand the emotional reaction as a reaction (perhaps socially or educationally mediated) to the insight into the rightness of that position.

We can also turn this point around. If we can show, through a convincing argument, that Kantianism is false, we have no need for any empirical evidence to back this up further. Empirical evidence does not and cannot make Kantianism any more false if there already is a convincing argument that shows that it is false. Furthermore, if we do not yet know whether the claims of Kantian moral theory are true or false, empirical facts do not provide decisive proof one way or another.

In other words, if we can show, with good arguments, that Kantianism is true, empirical evidence does not debunk it; if we can show with good arguments that it is false, debunking it with empirical evidence is neither necessary nor helpful for philosophical purposes. The factual premises of the debunking argument may serve some other purpose, such as curing us of recalcitrant emotions, but that is a psychological follow-up task that presupposes that the philosophical work has already been done. By themselves, the empirical premises of Greene's debunking argument do not establish the truth or falsity of the core tenets of Kantian ethics.

The most the empirical evidence can do in the context of normative ethics – and this may be important enough – is to prompt us to *think something through* that has wrongly been taken for granted. Suppose you have been made to believe, as the result of hypnosis, that 12,345 +

67,890 is 80,235. Suppose too that you did not know this before the hypnosis. If you are later told that you came to believe this merely as the result of hypnosis, should you now believe it is false? Of course not. Your having been led to believe it through hypnosis is compatible with its being true. The revelation does, however, give you good reason to suspend the belief and either remain agnostic or do the calculation. By analogy, the same is true of facts about human emotional tendencies and their bearing on moral judgment. Given that Kantians ground their arguments in core assumptions about reason and valid reasoning, arguments to the effect that the moral judgments of test subjects reflect emotional reactions (which in turn might be explained in terms of evolutionary theory or human psychology) do not by themselves show that Kantianism is wrong. Kantian theorists will argue that the circumstances under which a judgment is formed do not necessarily affect the validity of the judgment as such, but that knowledge of these circumstances may well constitute good reason to pause and examine the justification for one's position very critically.

In other words, the third premise of Greene's debunking argument is indeed crucial to the success of his argument as a whole. In the strong version required for Greene's strong conclusion, the premise builds in the claim that Kantianism is not supported by convincing argument, thereby presupposing what the argument is supposed to establish. If we remove the third premise and allow for the possibility that Kantian ethics can or might be justified through argument, the remaining premises do not lead to the conclusion that Kantian ethics is a form of moral confabulation.

This point finds confirmation in Greene's recent reformulation of his argument in response to Selim Berker's criticisms (Greene (2010)), because this new version has the same structural flaw. Greene now introduces the example of the incest taboo. Science tells us that the repulsion many people feel for consensual adult incest derives from a biological adaptation that avoids genetic diseases, Greene argues, claiming that once we realize this and think clearly, we will no longer see a need to maintain the incest prohibition in the case of consenting adults who use birth control (Greene (2010)). He summarizes his 'normative conclusion' as follows: 'Insofar as consensual adult incest is not on the whole harmful, and insofar as we lack a non-intuition-based justification for condemning consensual adult incest, we have no reason to believe it is wrong' (Greene (2010), 10). Greene discusses the incest example to illustrate his general claim that scientific information contributes to the debunking of deontology in general and Kantianism in particular (Greene (2010)).



Greene's incest example again makes very clear, however, why the empirical information does not debunk Kantianism. In this example, the crucial evaluative premise is that *there is no rational justification* for condemning consensual adult incest (in Greene's words, the premise is that 'we lack a non-intuition-based justification for condemning' it). It is *only* because of this premise that the evolutionary explanation of the incest taboo, together with the premise that consensual adult incest is not harmful, leads to his conclusion that we have no reason to believe that consensual adult incest is wrong. If we bracket that evaluative premise, the evolutionary account of the incest taboo tells us *nothing* about the moral permissibility of incest. This is easy to see when we consider the case of incest with children. Assuming that there are strong moral *reasons* for condemning incest with children, the fact that the incest taboo can be explained in evolutionary terms makes no moral difference whatsoever. Conversely, if we add Greene's premise that consensual adult incest cannot be condemned on rational grounds, there is nothing left to debunk. At most, evolutionary causal explanations could play a therapeutic role in curing us of recalcitrant emotions, but their role would not be justificatory, and the question whether certain emotions are 'recalcitrant' or, rather, morally helpful has to be settled independently. In sum, if the incest prohibition has a convincing rational justification, the causal explanation will not debunk it; and if it can be shown to be unjustified, there is nothing left to debunk.[4]

Most Kantians are not going to grant Greene the premise that Kantian ethics does not have a convincing justification. Greene, however, entirely sidesteps any discussion of the relative merits of the *arguments* in favor of Kantianism. He would first need to argue that the justifications of Kantianism fail, before his debunking argument gains any traction. He simply *assumes*, however, that Kantianism is ill-founded (that 'we lack a non-intuition-based justification' for it). This assumption would need to be proven; it cannot simply be taken for granted. Moreover, if it can be proven, we do not need any further 'debunking'.

In sum, Greene's 'debunking' argument against Kantian ethics, grounded in his neuropsychological research concerning the role of emotions in the empirical formation of moral judgments, does not succeed. In the absence of the evaluative premise, the factual premises do not prove anything concerning the justification of Kantian moral rationalism; but in order to justify the inclusion of the evaluative premise, one has to do the hard work of arguing, and once one has succeeded in securing the evaluative premise, debunking is no longer necessary. At

least in the case of Kantian ethics, one cannot – as Greene puts it – 'spot a rationalizer without picking apart the rationalizer's reasoning'.

## 2 It's just wrong!

Given their argumentative weakness, why do such empirical debunking arguments against Kantian ethics have any initial plausibility at all, in the eyes of a considerable number of people? I believe that the answer can be found in the widely shared understanding of Kant and Kantianism as merely issuing pronouncements that some actions 'are just wrong!' or 'simply must be done'. On this view, Kantianism resembles a form of direct moral intuitionism that somehow understands its core intuition as rational. If Kantian ethics is interpreted along such direct-intuitionist lines, it is likely to be regarded as a form of verbal foot-stamping or fist-thumping, with no convincing argument from which its pronouncements follow: You just ought never to lie! It's just wrong! This is how Greene portrays Kant, as we have seen above. On this understanding of the Kantian project, if its alleged core intuition is 'unmasked' as an *emotional* response, then this might seem to imply that we do indeed 'lack a non-intuition based justification' for Kantianism.

The direct-intuitionist understanding of Kantianism on which this inference is based, however, is a misrepresentation, and this is why the unmasking argument misses its intended target. In fact, Kantians usually argue strongly *against* rational intuitionism, and intuitionism runs counter to the Kantian understanding of autonomy of the will. Any time moral values or principles are grounded in something other than the will itself – such as human nature, tradition, or an independent realm of moral truth to which we have access through moral sense or rational intuition – heteronomy results. Accordingly, many current Kantian ethicists are opponents of metaethical substantive realism. They explicitly deny the claim that moral values can be grasped by rational intuition (Korsgaard (2009), 64–5; O'Neill (1989), 206–18). If they do use the label 'realism' to describe their own view at all, it is qualified so as to indicate the difference from substantive realism (such as the *procedural* realism defined by Korsgaard as the view 'that there are answers to moral questions; that is, that there are right and wrong ways to answer them' (Korsgaard (1996a), 35).

Common caricatures notwithstanding, Kantian theorists typically believe they have an *argument* regarding which action principles are morally justified and which are not, and why. They do not typically simply appeal to some direct intuition that a certain action or maxim of



action 'is just wrong' or 'is just right'. Instead, in distinguishing sharply between questions of moral justification and questions of empirical judgment formation, many Kantian ethicists argue that there are *good reasons why* specific moral commands are what they are, even if ordinary people in ordinary life do not always have these reasons clearly before their eyes. These reasons are grounded in the structure of agency or the practice of reasoning about action. The idea is typically that there are certain rational commitments that we undertake when engaging in acting or reasoning about action, and that these rational commitments entail certain conclusions regarding moral issues.[5] As Onora O'Neill puts it, the idea is to 'use minimal and plausible assumptions about human rationality and agency to construct an account of ethical requirements that is rich and strong enough to guide action and reflection' (O'Neill (1989), 194).

There are many varieties of Kantianism, so the way this gets spelled out varies greatly. Usually, however, Kantian ethicists formulate the rational criterion in terms of notions such as the autonomy of the will, consistency in action, the 'universalizability' of maxims, or the nature of reasons. In most cases they focus on the question of what can be rationally defended as good action. Despite the great variety of approaches among Kantian ethicists, they generally argue in terms of a conception of consistency, a theory of practical reasoning, a philosophy of action, and so on. They do not, at least not typically, merely perseverate[6] in assertions that some action or action principle 'is just wrong' or 'is just right'. Nor do they proceed on the basis of the empirical assertion that most people reach their moral judgments by conscious reasoning or that most people act rationally most of the time.

Kantians need not disagree that 'natural' emotional reactions may influence moral judgment formation, but they will insist that questions concerning the empirical genesis of moral judgments should be distinguished clearly from questions concerning their normative validity, and that the former cannot by themselves undermine the latter. Again: if there is a rational justification for Kantianism, empirical psychology does not debunk it; and if we know that Kantian ethics cannot be rationally justified, empirical evidence concerning the role of emotions in the genesis of moral judgment formation has nothing to debunk.[7]

## 3 Following Kant's lead: the moral importance of psychological research

The fact that empirical psychology does not debunk Kantian ethics does not mean that empirical psychology holds no interest for moral theory

and practice. On the contrary, as I argue in this section, on Kantian grounds it is arguably even an indirect *duty* to take at least some interest in morally relevant empirical psychology.

As I mentioned, current Kantian ethicists tend not to pay much attention to empirical research in moral psychology.[8] This may reflect a traditional tendency in moral theory more generally, rather than anything specific to Kantians, or it may have to do with a specific tendency among Kantians to focus more on the justification of principles than their application. However that may be, Kantian ethics can easily accommodate empirical psychology and use its results productively, and I shall argue that it should.

Strikingly, Kant himself agreed with current psychologists who deny that agents are fully transparent to themselves, and who claim that humans tend to ascribe moral motives to themselves even when they act from self-love. Kant himself had no illusions about human weaknesses or the extent of unconscious decision-making that is guided by feelings instead of reasons. He made a point of arguing that empirical observation of human agents leads one to doubt whether genuine virtue exists anywhere at all. He wrote in the *Groundwork*, for example, that

> it cannot be inferred with certainty that no covert impulse of self-love, under the mere pretense of that idea [*viz.*, of duty], was not actually the real determining cause of the will; for we like to flatter ourselves by falsely attributing to ourselves a nobler motive, whereas in fact we can never, even by the most strenuous self-examination, get entirely behind our covert incentives.... (G 4:407)

Rather than conceding that this insight debunks his moral theory or undermines moral practice, however, Kant claimed that knowledge of one's pre-reflective impulses and emotional responses should be used to *enhance* effective moral reasoning and action. It may help us become aware of our own and others' biases and other obstacles to acting morally, as well as provide us with strategies to overcome them effectively. Here, empirical psychology can provide invaluable information. For this reason, Kant was actively interested in the moral relevance of empirical psychology ('moral anthropology'). He regarded this as a necessary part of practical philosophy, calling it 'indispensable' (MS 6:217).

For Kant, the point of gaining knowledge about empirical human psychology is to provide an account of 'the subjective conditions in human nature that hinder people or help them in *fulfilling* the laws of a metaphysics of morals' (MS 6:217). This knowledge is to be conducive



to the 'development, spreading, and strengthening of moral principles (in child-rearing, in school education, and in popular instruction)' (MS 6:217). Kant's aim was to put the relevant empirical knowledge at the service of moral agency. He explained this relationship by drawing an analogy with memory enhancement. Knowledge of the mechanisms involved can be used to improve one's memory; if you know what helps you memorize certain things more easily, you can adjust your memorizing strategies. Similarly, knowledge of empirical psychological features and processes can be put to use in the service of our practical goals, including our moral ones (A 7:119). Kant discusses human emotions, for instance, arguing that we more easily fulfill our moral duty of beneficence when we cultivate our naturally compassionate emotional responses. Visiting hospitals or debtor prisons tends to trigger compassionate feelings, he writes, and we can put this emotional reaction to good moral use in service to fulfilling our duty of beneficence (MS 6:457; *cf.* also MS 6:456).

Kant even calls it a *duty* to cultivate certain emotional responses and to make use of those that are already available and in outward agreement with what duty demands of us.[9] He writes that it is an

> *indirect* duty to cultivate the compassionate natural (aesthetic) feelings in us, and to make use of them as so many means to sympathy based on moral principles and the feeling appropriate to them. (MS 6:457)

An indirect duty is a duty in the service of morality, and Kant does indeed speak of 'using' feelings as 'means'. The general idea here seems to be that if there are means that may be used to reduce obstacles to moral agency or that may enhance its effectiveness, then a moral agent cannot rationally will, as a universal law, the maxim to ignore them. Thus, a moral agent who has adopted the maxim to help others in need, from duty, cannot consistently will a maxim to leave unconsidered the available knowledge regarding ways to attain the goal of helping others in need. Given that agents will to accomplish moral goals, it would be irrational for the agent to adopt a maxim to ignore information about hindering or helpful psychological factors relevant to accomplishing one's goals if such information is available. This argument amounts to an argument in support of the (imperfect) duty to pay some attention to morally-relevant empirical psychological knowledge, in the service of moral goals. Being an imperfect duty, this duty does not specify exactly what or how much one should do in this regard. This is up to the agent,

as is the case with other imperfect duties such as the duty of beneficence. But one ought not to neglect empirical psychology as a matter of principle (*i.e.*, on a maxim of neglecting it), for example by mistakenly assuming that the non-empirical grounding of Kantian ethics makes empirical psychology altogether irrelevant to moral agency.

As is clear from this argument, the indirect duty to use empirical psychology in the service of morality *presupposes* a basic moral principle. First, within Kantian ethics empirical psychology does not play a role at the level of the justification of the basic moral principle. As Kant put it, such empirical knowledge 'must not precede a metaphysics of morals or be mixed with it' (MS 6:217). Once the derivation and justification of the basic moral principle is given, though, the question emerges which factors create or reduce obstacles to moral agency and to its efficacy. Kant recognized this as crucial for educational and instructional purposes, not just for children but for any moral agent.

Second, it is important, from a Kantian perspective, to emphasize that the use of natural emotional response tendencies for moral purposes, and the express cultivation of morality-supporting emotional responses, *presuppose* moral agents and are undertaken *in the service of* moral agency. They are not meant to serve as a non-moral substitute for moral agency, nor do they render moral agency superfluous. Agents who visit hospitals so as to trigger sympathetic feelings in themselves are doing so in the service of their already existing moral goals. They are not relinquishing their moral agency to an autopilot.

This point requires some elaboration, however, because it might seem to some as though using emotions in this way would 'outsource' moral agency to one's natural responses and hence would constitute an unacceptable form of moral evasion or indolence. In the case of the hospital and prison example, it might seem as though the agent is slacking: it might seem as if the agent takes a shortcut through natural psychological mechanisms, rather than doing what is right 'from duty'. This agent might seem to be the moral equivalent of a marathon runner who, in order to reach the finish more easily, hitchhikes over a difficult part of the course. The criticism that Patrick Frierson articulated, from a Kantian perspective, of the 'situation management' recommended by 'situationist' moral psychologists, might then seem to be appropriately directed at Kant's own proposal, namely, that such a strategy leads agents to 'preserve corrupt volitional structures while becoming increasingly morally self-satisfied' (*cf.* Frierson (2010a), 37).

In response to this objection, it is important to point out that a good moral agent who employs and cultivates his own emotional response tendencies is and remains driven by moral considerations. In other words, it is *presupposed* that this agent's volitional structure is not corrupt, and the short-cut is taken 'from duty'. The comparison with the marathon runner turns out to be inapt, because there is no additional moral requirement that is analogous to the rule governing marathons, namely, that you reach the finish on foot and *only* on foot. Certainly, moral agents ought to act *from duty*, but nothing in Kantianism requires them to do so *without* any sentiment, let alone to do so without any supportive sentiments. If morality requires you to promote a certain moral goal (such as the well-being of others), and you have already set yourself this goal from duty, there is no additional requirement for you to reach it without the help of supportive sentiments; and if you do employ emotional mechanisms in order to reach the moral goal, again you do *that* from duty, too. In other words, if your volitional structure is morally good, the morally motivated use of your emotions in the service of your moral goals fits within that very same volitional structure.



Some Kantians might object (to Kant) that using knowledge of one's psychological tendencies in the service of morality introduces *auxiliary motives*, for it might seem that the psychological impulse serves as a second motive *next to* the moral one. This worry is unnecessary, however. If the moral motive is what drives the action, the psychological impulse does not have equal motivational status. If an agent acts from duty, the moral motive is sufficient and remains 'in the lead', and the helpful psychological tendency is enlisted in service of attaining the moral goal. This means that whatever help is sought from psychological tendencies, such tendencies are subordinated to and governed by the motive of duty.

I have used Kant's example of visiting hospitals and prisons for the sake of beneficent action, but he used this as an example of the more general point that we have an indirect duty to cultivate morally supportive natural feelings. Kant also mentions other examples of indirect duties that are similarly based on knowledge of human psychology. The best known of these is probably his claim, in the *Groundwork*, that it is an 'indirect duty' to secure one's own well-being as a way of reducing one's susceptibility to temptation (G 4:399). Here, too, Kant argues that knowledge of our psychological tendencies needs to be taken seriously, for the sake of morality, and that it has implications for how we ought to act. It informs our conception of what our duties are: not at the level of the formulation of the basic moral principle itself, but at the level of its application to human beings.

On similar grounds, current Kantian ethicists can and should be wholeheartedly interested in empirical moral psychology, for the sake of moral agency. Moreover, whereas Kant was restricted to eighteenth-century empirical psychology, largely the result of armchair theorizing and personal observation, today's moral theorists have better research results at their disposal.

To give one concrete example of more recent psychological research that may become positively helpful to Kantian moral agents (or any agent, for that matter), let me point to psychological research on self-regulation. This shows, among other things, that setting oneself a general goal and then simply trying very hard to achieve it is generally not the most effective approach. As Peter Gollwitzer has documented, the additional adoption of 'implementation intentions', spelling out in advance the 'when, where, and how' of goal-striving, leads to a much higher success rate than merely intending the goal. This is the difference between 'When I eat dinner, I will drink water, not beer, in order to lose weight' and 'I intend to lose weight'. Merely setting a goal does have a noticeable effect, but the mere goal-setting, even with very 'strong' intentions, is not nearly as effective as adding to this goal an implementation intention in the form of a specific if-then plan of action (see the meta-analysis in Gollwitzer and Sheeran (2006) and also Schweiger Gallo *et al.* (2009)). This has been shown to hold for goals ranging from emotion regulation (e.g., the goal of reducing one's own reactions of disgust or fear) to altering one's behavior or accomplishing goals in the external world. It has also been shown to reduce obstacles on the road to goal achievement, ranging from trouble getting started to derailments (distractions, temptations, *etc.*), and internal interferences (anxiety, disgust, exhaustion, overconfidence, *etc.*).

Such findings have obvious relevance to Kantian ethics. To mention just one example, imperfect duties such as 'promote the well-being of others' are usually formulated as 'goal intentions' only, and agents should be aware of the importance of formulating 'implementation intentions' concerning the specific ways in which they intend to help in practice. Rather than only rehearsing a general goal intention, say, to 'be more beneficent in the New Year', they would do well to make their aims more specific by specifying the activities they aim to undertake and the moments when they plan to do so.

Of course there are bootstrapping problems here – how do I ensure that I form implementation intentions? – and all sorts of other limitations that will continue to interfere with our goal attainment. The important point, though, is that we can learn to replace worse strategies with better ones, and that using better strategies has a 'medium-to-large effect' (Gollwitzer and Sheeran (2006)) on our success in attaining our goals, including our moral ones.



This is merely one example, but of course the list of relevant research results is long. It includes work on implicit biases, stereotypes, order effects, framing effects, priming, and so on – work that identifies influences of various sorts that may interfere with one's moral agency. It also includes work on effective strategies, including social 'scaffolding' conditions, for efforts at self-regulation and self-correction to have a real-life effect.

Clearly, Kantians – just as much as anyone else – have good reason to take note of this and other work in empirical psychology, with an eye to both the conditions related to *setting* moral goals and those related to the effectiveness of *attaining* them. It helps to become aware of obstacles to doing what morality requires, and to develop effective strategies to overcome them. As mentioned above, Kant himself regarded such knowledge as 'indispensable' (MS 6:217), and current Kantian moral theorists have good reason to follow his lead.

## Conclusion

In sum, the significance of empirical psychology for Kantian ethics lies not in its potential to show that Kantianism is mistaken. Greene's debunking strategy fails because it begs the question. Furthermore, Kant and Kantians are well aware of the fact that humans have a tendency to dissemble and rationalize. Their moral theory is grounded not in an overly sanguine view of human nature, but in presuppositions that are – or so they argue – always already implicit in practical reasoning and agency. This grounding enables Kantians to distinguish between the empirical genesis and the normative validity of moral judgments, and to claim that their moral theory does not rest on assumptions about the former.

Nevertheless, Kantian moral theorists should be interested in empirical psychological research bearing on moral agency. On Kantian grounds, it is an indirect and 'imperfect' duty to acquaint oneself with the psychological conditions that hinder or support one's attempts to act morally and reach one's moral goals, in order that this knowledge can inform one's moral agency. In this context, empirical psychology has a much more significant role to play in Kantian ethics than is commonly assumed.[10]

## Notes

1. See, for example, Merritt *et al.* (2010) and Doris (2002).
2. 'Emotion', in this context, refers to emotions grounded entirely in natural psychological processes; Greene does not consider the feeling of respect in this context, and for the sake of argument I shall bracket it as well. Also, there are other forms of deontology besides Kantianism, but for the purposes of this chapter I focus only on Kantian ethics.
3. A re-analysis of the Greene *et al.* data revealed that a few dilemmas had extremely large effects and that this skewed the averages. A non-trivial number of cases (9 of the 40 moral dilemmas, 8 of which were in the 'personal moral' category) were answered almost unanimously and very quickly, the dominant answer being scored as deontological. To give one example, this was the case for the 'dilemma' called the 'hired rapist', where the question was whether it would be appropriate for a husband to hire a rapist to rape his wife, so that he could comfort her afterwards and she would appreciate him more. Test subjects were fast and practically unanimous in judging this to be inappropriate. But Greene averaged such results with the response time to dilemmas that received more varied responses (such as the footbridge trolley dilemma), and this made it seem as if deontological answers were on the whole faster than consequentialist ones. In McGuire's re-analysis, if one brackets the cases on which there was more than 95 percent agreement among test subjects (such as the hired rapist case), deontological and consequentialist judgments took *equally long*. Greene's statistical result was entirely due to the group of 'non-dilemmas' such as that of the hired rapist. Selim Berker also points to this problem (Berker (2009), 308–11).
4. There is of course much more to say about evolutionary debunking arguments, but note that such arguments do not touch on Kantianism in the same way as they touch on moral realism, insofar as the latter is understood as the view that there are 'mind-and-language independent' moral truths (see Clarke-Doane (2012)). Kantians typically do not regard moral principles to be 'mind-and-language independent', because they regard them as grounded in reason. Reason (which is central to Kantian ethics) can be understood as a 'third-factor explanation' to explain why it is not a mere coincidence that the moral principles that are valid are also believed to be valid. For a recent critique of evolutionary debunking strategies and the possibility of 'third-factor explanations', see Wielenberg (2010).
5. There are notable exceptions, for example, Wood (2008).
6. Perseveration is a term used in psychology to describe the tendency to repeat a particular response even after the initial stimulus has ceased, or the inability to change one's behavior in the light of changed circumstances or information.
7. For a critical discussion of the methodology used in much recent social psychology research on moral judgment (e.g., in the work of Haidt), see Kennett (2012).
8. There is no discussion of it, for example, in the work of Christine Korsgaard, Onora O'Neill, Thomas Hill, or Allen Wood, even though their writings include section titles such as 'Problems of bringing the kingdom down to earth' (Hill (2000), 51–5), 'The psychology of action' (Korsgaard (2009), 104–8, in a chapter on 'Autonomy and efficacy'), 'Embodied obligations' (O'Neill (1996a), 146–53), or 'Human nature' (Wood (2008), 4–6). Even Barbara Herman, who comes closest and aims to 'let the phenomena in' by paying attention to 'what we are like as agents' (Herman (2007), vii), does not address the question of the importance of recent empirical psychological research for moral theory and practice. Within Kant scholarship, there is growing interest in the role of emotions in Kant's work, as is evidenced by the present volume.
9. See also Sherman (1990) and Borges (2008).
10. Work on this chapter was partly funded by the Netherlands Organization for Scientific Research (NWO).

**Bibliography**


Berker, Selim (2009). "The Normative Insignificance of Neuroscience," *Philosophy and Public Affairs* 37: 293-329.

Borges, Maria (2008). "Physiology and the Controlling of Affects in Kant's Philosophy," *Kantian Review* 13: 46-66.

Clarke-Doane, Justin (2012). "Morality and Mathematics: The Evolutionary Challenge," *Ethics* 122: 313-340.

Dean, Richard (2010). "Does Neuroscience Undermine Deontological Theory?" *Neuroethics* 3:43-60.

Doris, John M (2002). *Lack of Character: Personality and Moral Behavior*. Cambridge: Cambridge University Press.

Frierson, Patrick (2010). "Kantian Moral Pessimism," in *Kant's Anatomy of Evil*, eds. P. Muchnik and S. Anderson-Gold. Cambridge: Cambridge University Press, 33-56.

Galvin, Richard (2011). "Rounding Up the Usual Suspects: Varieties of Kantian Constructivism in Ethics," *Philosophical Review* 61: 16-36.

Gollwitzer, Peter M. and Paschal Sheeran (2006). "Implementation Intentions and Goal Achievement: A Meta-Analysis of Effects and Processes," *Advances in Experimental Social Psychology* 38:69-119.

Greene, Joshua (2008). "The Secret Joke of Kant's Soul," in *Moral Psychology*, ed. W. Sinnott-Armstrong. Cambridge, MA: MIT Press, vol. 3, 35-79.

Greene, Joshua (2009). "Dual-Process Morality and the Personal/Impersonal Distinction: A reply to McGuire, Langdon, Coltheart, and Mackenzie," *Journal of Experimental Social Psychology* 45: 581-584.

Greene, Joshua (2010). "Notes on 'The Normative Insignificance of Neuroscience' by Selim Berker," http://www.wjh.harvard.edu/~jgreene/GreeneWJH/Greene-Notes-on-Berker-Nov10.pdf. Last accessed January 29, 2014.

Greene, Joshua D., Fiery A. Cushman, Lisa E. Stewart, Kelly Lowenberg, Leigh E. Nystrom, and Jonathan D. Cohen (2009). "Pushing Moral Buttons: The Interaction between Personal Force and Intention in Moral Judgment, *Cognition* 111: 364-371.

Greene, Joshua D., Sylvia A. Morelli, Kelly Lowenberg, Leigh E. Nystrom, and Jonathan D. Cohen (2008). "Cognitive Load Selectively Interferes with Utilitarian Moral Judgment," *Cognition* 107: 1144-1154.

Greene, Joshua D., R. Brian Sommerville, Leigh E. Nystrom, John M. Darley, and Jonathan D. Cohen (2001). "An fMRI Investigation of Emotional Engagement in Moral Judgment," *Science* 293: 2105-2108. Supplementary material: www.sciencemag.org/cgi/content/full/293/5537/2105/DC1.

Haidt, Jonathan (2001). "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* 108:814-834.

Herman, Barbara (2007). *Moral Literacy*. Cambridge, MA: Harvard University Press.

Hill, Thomas E., Jr. (2000). *Respect, Pluralism, and Justice: Kantian Perspectives*. Oxford: Oxford University Press.

Kahane, Guy (2011). "Evolutionary Debunking Arguments," *Noûs* 45: 103-125.

Kennett, Jeanette (2012). "Living with One's Choices: Moral Reasoning *in vitro* and *in vivo*", in *Emotions, Imagination, and Moral Reasoning*, eds. R. Langdon and C. MacKenzie. New York: Psychology Press, 257-278.

Korsgaard, Christine M. (1996). *The Sources of Normativity*. Cambridge: Cambridge University Press.

Korsgaard, Christine M. (2009). *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

McGuire, Jonathan, Robyn Langdon, Max Coltheart, Catriona Mackenzie (2009). "A Reanalysis of the Personal/Impersonal Distinction in Moral Psychology Research", *Journal of Experimental Social Psychology* 45: 577-580.

Merritt, Maria W., John M. Doris, and Gilbert Harman (2010). "Character," in *Moral Psychology Handbook*, ed. Walter Sinnott-Armstrong. Oxford: Oxford University Press, 355-401.

Nisbett, Richard E. and Timothy D. Wilson (1977). "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review* 84: 231-259.

O'Neill, Onora (1989). *Constructions of Reason: Explorations of Kant's Practical Philosophy*. Cambridge: Cambridge University Press.

O'Neill, Onora (1996). *Towards Justics and Virtue: A Constructive Account of Practical Reasoning*. Cambridge: Cambridge University Press.

Sauer, Hanno (2012a). "Educated Intuitions: Automaticity and Rationality in Moral Judgment," *Philosophical Explorations* 15: 255-275.

Sauer, Hanno (2012b). "Morally Irrelevant Factors: What's Left of the Dual-Process Model of Moral Cognition?" *Philosophical Psychology* 25: 783-811.

Schweiger Gallo, Inge, Andreas Keil, Kathleen C. McCulloch, Brigitte Rockstroh, Peter M. Gollwitzer (2009). "Strategic Automation of Emotion Regulation," *Journal of Personality and Social Psychology* 96: 11-31.

Sherman, Nancy (1990). "The Place of Emotions in Kantian Morality," in *Identity, Character and Morality*, eds. Owen Flanagan and Amélie Rorty. Cambridge: MIT Press, 149-170.

Wielenberg, Erik J. (2010). "On the Evolutionary Debunking of Morality," *Ethics* 120: 441-464.

Wilson, Timothy D. (2002). *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, MA: Harvard University Press.

Wood, Allen (2008). *Kantian Ethics*. Cambridge: Cambridge University Press.


Note: This bibliography is a selection from the general bibliography at the end of the volume; it is not part of the essay page range.