Causation in Moral Judgment

Michael D. Kurak (mdkurak@gmail.com)
Forthcoming in *Mind and Matter*

Abstract

Research on moral judgment is refueling public interest in an old debate concerning the general foundation of morals. Are moral judgments based on reason or on feeling? Recent research in moral psychology and neuroscience concludes that moral judgments occur rapidly, automatically, and largely without the aid of inference. Such findings are utilized to criticize moral theories that require deliberation to precede moral judgment as its cause. The main targets of this criticism are the moral theories of Piaget and Kohlberg, but Immanuel Kant's moral philosophy is also criticized for this failing. This essay defends Kant from this charge by clarifying the role of deliberation in his moral theory and by demonstrating that, for Kant, moral judgment invokes a self-organizing system that has the capacity to rapidly determine the moral permissibility of any desired purpose – real or imagined. Further support for Kant's moral theory is gleaned from recent work on self-organization in the brain.

1. Introduction

Although the origins of the debate on moral foundations predate the writings of David Hume, perhaps no one frames the problem as clearly. In his *Enquiry Concerning Principles of Morals* Hume discusses a controversy that has started of late, "concerning the general foundation of Morals; whether they be derived from Reason, or from Sentiment; whether we attain knowledge of them by a chain of argument and induction, or by an immediate feeling and finer internal sense;" whether, like all sound judgment concerning matters that will admit of truth and falsehood, [moral knowledge] is "the same to every rational intelligent being; or whether, like

the perception of beauty and deformity," morals are "founded entirely on the particular fabric and constitution of the human species" (Hume 1751, §1). In this well-known passage Hume establishes a dichotomy in which morality is either based on "pure reason," in which case it is as invariant across the universe as physics, or it is based on a property peculiar to the human species, in which case its universality is restricted to human beings. Hume takes the latter position.

Hume's position has lately received considerable empirical support. Research has revealed that moral judgment takes place rapidly, automatically and largely without reliance upon inference (Haidt 2001, Bargh and Ferguson 2000, Murphy *et al.* 2009).[1] Such findings have been utilized by empirically-minded moral theorists to support a "model of moral judgment [that] builds on the insights of Hume to suggest that moral judgment is generally a result of quick gut feelings, much like aesthetic judgment" (Schnall *et al.* 2008). This model maintains "that moral intuitions (including moral emotions) come first and directly cause moral judgments" (Haidt 2001, p. 815). As a result, its advocates are critical of moral theories that place reason at the basis of moral judgment and that do not assign a substantial causal role to the emotions.[2]

Within moral psychology the main targets of this line of criticism are the moral theories of Piaget and Kohlberg (see Haidt and Bjorklund 2007). Piaget and Kohlberg are identified as being committed to the idea, which they appear to derive from Kant, that moral reasoning precedes moral judgment as its causal basis. Since the recent findings of empirical psychology

[1] This brief list is not exhaustive. There are various psychologized moral theories which share the automaticity thesis. For a review see D.K. Lapsley and P.L. Hill, 2008, On dual processing and heuristic approaches to moral cognition, Journal of Moral Education, Vol. 37, No. 3, September 2008, pp. 313-332.
[2] In an interview Haidt describes Kant's moral philosophy as "weird," "hyper-systematic," and overly rational. http://edge.org/3rd_culture/morality10/morality.haidt.html

[2] In an interview Haidt describes Kant's moral philosophy as "weird," "hyper-systematic," and overly rational. http://edge.org/3rd_culture/morality10/morality.haidt.html

are inconsistent with this idea, Paiget and Kohlberg are criticized for their theoretical commitment to it. Since Kant is also commonly understood as being committed to this "rationalistic" idea, his moral theory is also criticized as being inconsistent with recent empirical findings on moral judgment (see Greene 2007). In the case of Kant, however, the target of such criticism is a straw man. Kant's moral theory is not committed to the view that moral reasoning precedes moral judgment as its causal basis. Moreover, Kant's moral theory can readily accommodate the recent empirical findings concerning the automaticity and rapidity of moral judgments.

The essay proceeds first by briefly reviewing Hume's moral theory and by examining his reliance upon a particular model of perception. In the second section we review Kant's main objection to Hume's moral theory and we explicate his alternative account of the role of emotions in moral judgment. In section three we sketch two competing models of perception; one passive and based on the idea of transduction, and the other active and based on the idea of self-organization. We argue that while the moral theories of Hume and his contemporary representatives rely upon the idea of transduction, Kant's moral theory relies upon the idea of self-organization. Since such an interpretation of Kant's moral theory is novel, we proceed to a defense of it.[3] This defense demonstrates the ability of his theory to accommodate the recent empirical findings of moral psychology regarding the rapidity and automaticity of moral judgments.

2. The Empiricist Position

[3] The body of literature on Kant is enormous. I am unable to discover any previous interpretations of his moral philosophy that conceive of it in the manner that I do herein.

Hume's model of moral judgment is closely based on his understanding of aesthetic judgment which, in turn, is based on his limited understanding of perceptual judgment. Hume holds roughly to the view that perception occurs when an external object strikes an organ of sense, thereby agitating it and producing, as a result, an "impression" in the mind. This impression forms the basis of judgments of the properties of objects and of the experience of pleasure and pain. These judgments and experiences are informed by previous experiences with similar impressions. Thus, for example, light reflected from a painting of flowers agitates the organ of sight which produces an impression of the painting in the mind. The judgment that the painting is a painting of flowers is influenced by ones previous experience with flowers (see Hume 1739, T. 1.1 §§1 and 2).

For Hume, aesthetic judgment differs from straightforwardly perceptual judgment primarily by the nature the product of the agitated sense.[4] The judgment that a wine is red is based upon impressions which are the product of the agitation of all, or some of, the organs of sight, smell and taste. The judgment that this particular red wine is good requires subjecting these impressions to an internal aesthetic sense which, in response to this agitation, produces feelings of pleasure or displeasure. These feelings form the basis of the judgment of the quality of the wine.[5]

The moral judgment that one ought or ought not to drink the wine, perhaps because it is not mine, involves subjecting the impressions and facts under consideration to an internal moral

[4]Hume (1739) writes: "that beauty is nothing but a form, which produces pleasure, as deformity is a structure of parts, which conveys pain; and since the power of producing pain and pleasure make in this manner the essence of beauty and deformity, all the effects of these qualities must be deriv'd from the sensation..." (T. 2.1, §8).
[5]Sometimes it seems as if Hume wishes to argue that the experience of aesthetic pleasure or displeasure is a second order judgment on the impression and sometimes it seems as if that pleasure or displeasure is part of the judgment of the object. In either case it is clear that Hume maintains that all impressions generate feelings of pain or pleasure. "I believe it may safely be establish'd for a general maxim, that no object is presented to the senses, nor image form'd in the fancy, but what is accompany'd with some emotion or movement of spirits proportion'd to it" (Hume 1739, T. 2.2, §8).

sense, "or whatever you may please to call it," that produces feelings of approval or disapproval. These feelings form the basis of the judgment of the rightness or wrongness of the action under consideration. Thus, aesthetic and moral judgments are thought by Hume, on close analogy with perceptual judgments, to be based upon the product of the agitation of an internal sense, the operation of which is almost entirely mysterious.

Hume guarantees the possibility of inter-subjective agreement in moral judgment by arguing that nature has made the relevant internal (moral) sense "universal in the whole species" (Hume 1751, EPM §1). Reason's place in this scheme is to ensure that the relevant internal sense is provided with proper representations of the relevant facts and impressions. As Hume observes,

> …in many orders of beauty, particularly those of the finer arts, it is requisite to employ much reasoning, in order to feel the proper sentiment; and a false relish may frequently be corrected by argument and reflection. There are just grounds to conclude, that moral beauty partakes much of this latter species, and demands the assistance of our intellectual faculties, in order to give it a suitable influence on the human mind (1751, EPM §1).

Thus, aesthetic and moral judgment differ also from perceptual judgment in that the impressions and ideas that are subject to the internal sense can be modified by rational considerations. Be this as it may, the final arbiter of moral judgment is always the internal sense to which these facts and impressions are presented and which responds by producing feelings of approval or disapproval. As we will see below, some contemporary psychological accounts of moral judgment do not differ fundamentally from the account that Hume provides.

3. Kant's Position

As is well known, Hume's main opponent on the issue of moral foundations is Immanuel Kant. Kant's main objection to Hume's sensible proposal is that it is simply incapable of securing the possibility of morality at all. If one ought not to lie, cheat, or steal, for instance, then

this is to say that one must not do such things. On Hume's moral theory there is simply no way to provide "the ought" with the force of necessity that is required by the concept of duty. According to his theory, to say that one ought not to lie, cheat, or steal is only to say human beings are naturally disposed to disapprove of such actions; that, in general, their moral sense responds to such impressions and ideas with strong feelings of disapproval, which form the basis of moral judgments that such purposes ought not to be undertaken.

But feelings, observes Kant, can never provide the force of necessity that is demanded if morality is to be possible. Even if Hume is correct that we all share an internal sense that governs the way in which we respond to moral matters, this cannot account for duty, but only for inter-subjective disapproval. For Hume, actions are wrong because, even after considerable deliberation, we feel them to be wrong. If we cease to feel them to be wrong then they cease to be wrong. Consequently, for Hume, the force of duty is no more than the force of feelings. From considerations such as these Kant concludes that morality cannot be grounded in special properties of the human species (see Kant 1785, GMS 4:411 n.).

In order to highlight the difference between these positions we need only consider any morally laudatory action where we can predict the individual actors to have strong feelings that run counter to the performance of that action. We might consider, for example, the band members aboard the Titanic, who continue to play for some time after it becomes apparent that the ship is sinking. On Hume's account of moral foundations these individuals continue play because they have feelings of benevolence towards others that overcome the feelings that prompt them to save themselves.[6] For Kant, by contrast, the band members play on in recognition of a

[6] In fact, Hume would undoubtedly argue that the special circumstances of a sinking ship with insufficient lifeboats absolves them of all duty to continue to play (see Hume 1751, EPM 1 §3).

duty that is valid for them independently of feelings, but whose validity is subjectively indicated to them by feelings.

There are actually two closely related feelings at work on Kant's account.[7] The moral law generates (emotional) pain to the extent that feelings are present that incent one to actions that conflict with its commands. In our example, feelings of pain must exist to the extent that there are simultaneously feelings to flee or to do anything other than to remain and continue playing. Since people quite naturally want to save themselves from death, we can predict the existence of pain in such circumstances. However, from the perspective of reason, such feelings of pain have the effect of moving resistance out of the way. As Kant states, "in the judgment of reason this removal of an obstacle [e.g., feelings to flee] is esteemed equal to a positive furtherance of its causality" (Kant 1788, KpV 5:75).[8] Thus, we are humbled by the power of our own reason against our inclinations. The recognition of this power – the power of our true self – generates a positive feeling of respect for morality, as it manifests through us. This is *moral feeling* (Kant 1788, KpV 5:73). He characterizes it as "sublime" and "specifically different" from all other feelings. Moreover, he maintains "that the sublimity and inner dignity of the command in a duty is all the more manifest the fewer are the subjective causes in favor of it and the more there are against it" (Kant 1785, GMS 4:425). In this way, Kant renders intelligible the motivations of

[7] Although they may blend together in moral feeling, they are conceptually separable.

[8] The following footnote from the *Critique of Judgment* may be helpful in clarifying how pure practical reason, considered as a cause, can affect, and be affected by, natural causes. "One of the various supposed contradictions in the complete distinction of natural causality from the causality through freedom is given in the following objection to it. It is held that when I talk about nature putting *obstacles* in the way of the causality governed by laws of freedom (moral laws), or about nature *furthering* it, I do after all grant that nature *influences* freedom. But this is a misinterpretation, which is easily avoided merely by understanding what I have said. The resistance or furtherance is not between nature and freedom, but between nature as appearance and the effects of freedom as appearances in the world of sense; and even the causality of freedom (of pure and practical reason) is the causality of a natural cause (the subject, regarded as a human being and hence as an appearance) subject to [the laws of] nature. It is causality's determination whose basis is contained, in a way otherwise not explicable, in the intelligible that is thought of when we think freedom (just as in the case of the intelligible that is the supersensible substrate of nature) (Kant 1793, KU 196 n.).

someone who finds it within themselves to perform such heroic acts as those of the band members of the Titanic.

For both Hume and Kant emotions are crucially involved in moral judgment. The central philosophical issue concerns whether such feelings are the cause or the effect of moral judgment. On Hume's account, they are the cause. On Kant's account, they are the effect. For Kant, moral judgment depends only upon pure practical reason; that is, upon the power of reason apart from anything empirical, i.e., apart from feelings.[9] Kant argues that this is only way that we can secure the possibility of morality at all.

Of course, such a statement immediately brings to mind a deliberative act of reasoning (i.e., discursion). If it is true that "No man ought ever to lie" and also that "Socrates lied" then it will follow necessarily that Socrates did something that he ought not to have done. But, as we have noted, any contemporary theory of moral foundations must accommodate two well established features of moral judgment which, on the face, seem to preclude reason as a viable candidate for its basis; that is, the quickness with which moral judgments are made and their apparent automaticity. The moral theories of Piaget and Kohlberg are perhaps rightly criticized for not being able to accommodate such features. Kant's moral theory, as I will shortly demonstrate, is not dependent upon discursion in this manner and can accommodate these two features of moral judgment.

[9] Kant (1797) states that "every determination of the elective will proceeds from the idea of the possible action through the feeling of pleasure or displeasure in taking an interest in it or its effect to the deed; and here the sensitive state (the affection of the [so called] internal sense) is either a pathological or a moral feeling. The former is the feeling that precedes the idea of the law, the latter that which may follow it" (MST 400).

4. The Influence of Analogies with Perception

As we have seen, the idea of a moral sense is an idea born from an analogy with our five perceptual senses. Just as our perceptual senses provide fairly reliable information about the external world so, by analogy, does our moral sense. Just as our perceptual senses have no need of inferential processes to make judgments about the external world, so neither does our moral sense. A moral sense is, thus, conceived as a power of perceptual-like judgment that responds more or less automatically to representations of morally relevant information from the external world and, by this means, generates a product (i.e., a feeling) upon which to base judgments about that world.

There are, of course, obvious differences between our perceptual senses and that capacity for moral judgment which is being conceived as a moral sense. All of the perceptual senses are supported by corresponding sense organs, each of which is comprised, in part, of an immense number of receptors. The receptors of these sense organs are devoted to separate sensory modalities. There is, by contrast, evidence neither for the existence of a moral sense organ nor for the existence of a corresponding moral modality in the external world. Kant maintains that "we have no special sense for good and evil any more than for truth, although such expressions are often used" (Kant 1797, MST 400). Nevertheless, the analogy of moral judgment with aesthetic and perceptual judgment persists.[10]

The mainstream contemporary view of neuroscientists, psychologists and many informed lay persons is that the sense organs and their associated regions in the cortex are transducers of information from the external world. On this view, information is an inherent 'feature' of objects. The contours of a face, the phonemes of a word, the chemical constituents of an odor,

[10] Note that the classic sense of the term "aesthetic" denotes perception by the senses or by the mind, from the Greek *aisthetikos* "perceive."

are all examples of such object 'features.' A perceptual judgment involves assemblies of neurons in the cortex that match these features firing in response to them (see Singer and Gray 1995, Rizzolatti and Craighero 2004, Quian Quiroga and Panzeri 2009). This matching process is held by those cited to be fundamental to the perception of an object.

A moral judgment can be thought to involve a similar type of matching process in which assemblies of neurons fire in response to patterns of events, or morally salient features of the world, thereby generating a mental consequence, such as feelings of approval and disapproval. Thus, Johnathan Haidt, expresses an increasingly popular view when he writes that: "Human beings come equipped with an intuitive ethics, an innate preparedness to feel flashes of approval or disapproval toward certain patterns of events involving other human beings" (Haidt and Joseph 2004, p. 56).[11] According to Kant, theorists who maintain such a position "are audacious enough to turn a deaf ear" to "the voice of reason in reference to the will," perceptible even by "the commonest human being,...in order to uphold a theory that does not require them to rack their brains" (Kant 1788, KpV 5:35).

We can now observe that there is a competing view of the functioning of the perceptual senses that squares well with Kant's moral theory and that has the capacity to provide us with a competing model of moral judgment. On this view, instead of perceptual judgment being conceived passively, as a process which mechanically produces a response to an input, it is conceived actively, as a process in which the interaction of neurons gives rise to constraints which, at the same time, govern those very interactions. On this account, for each sense modality, sense perception is not held to involve small assemblies of feature-detector neurons firing in response to a given stimulus, but rather to involve large assemblies of neurons

[11] This view is shared by a fairly large number of individuals from various fields, including philosophy (see, for example, Damasio 1994, Haidt 2001, Cushman et al., 2006, Joshua Greene 2001, 2004, Sober and Wilson, 1998).

interdependently generating a spatial pattern (as an effect) which can, at the same time, be judged to be governing those very interactions (as a cause). This dynamic process of pattern formation in the brain is held to underpin perceptual judgment.

A central proponent of this view is the neuroscientist Walter J. Freeman. Freeman is best known for his work on rabbit olfaction and it is to this work that I turn here to explain the view that he advances. Freeman emphasizes that the neurons that comprise the olfactory bulb do not silently wait for input. Instead they show constant background activity which sensory input effectively disturbs. With each inhalation there is a burst of activity in the olfactory bulb that ends after exhalation. The burst manifests itself in the formation of a momentarily stable spatial pattern that incorporates the entire bulb. This pattern is, for any known odourant, consistent over repeated trials. Consequently, Freeman infers that such pattern formation is fundamental to object categorization, as it occurs in perception.

What is significant for our purposes is how this pattern is formed. First, every neuron of the olfactory bulb is involved. As Freeman (1999) observes, "every pattern must have dark as well as light and it is a mistake to assume that a neuron that is not firing in response to a stimulus is not part of a pattern; it may be silent because that is the role assigned to it by the macroscopic state of the bulb" (p. 100). Second, the pattern is not simply completed mechanically as a direct result of input from the stimulus; that is, the pattern is not a transduction of the stimulus input. Freeman writes:

> Before the state transition upon each inhalation, the bulbar neurons respond mostly to the input, but in the burst they respond mostly to each other. The bulb keeps an open door up to the state transition, but then the door is shut and the AM (amplitude modulation) pattern is determined by the synaptic connections in the bulbar neuropil, rather than by the stimulus (1999, pp. 101-2).

Since the door is shut to stimulus input during the state transition (i.e., during the burst, the state transition being, like a moral judgment, virtually instantaneous), the pattern formation is

governed by constraints that develop in the interaction of each neuron with all the other neurons that comprise the bulb.[12] Third, the resulting pattern is, at the same time, an "order parameter" of the system that "enslaves" the all neuronal elements of the olfactory bulb (Freeman 1999, pp. 111 and 114). In other words, the large-scale pattern that we conceive as an *effect* of the dynamic interactions of all the neurons of which the bulb is comprised, can simultaneously be conceived to be a governing *cause* of those same dynamic interactions. Fourth, the resulting system is one in which everything – the large-scale pattern that the bulb forms and the activity of each element of that pattern – can be said to produce everything else. Such a system is classically self-organizing.

Now, as we have noted, any understanding of moral judgment that depends upon an analogy with a model of perceptual judgment will be substantively affected by that model. If, on one hand, we employ a passive, empiricist model of sense perception, that relies upon the concept of transduction then the result will be a fundamentally mechanical model of moral judgment. If, on the other hand, we employ an active, self-organizing model of sense perception, such as the one advanced by Freeman, then the result will be a model of moral judgment that relies upon the idea of a self-organizing system. We can now ask the question: "What might such a model of moral judgment look like?" Although Kant's model of perception and the extent to which that model can reasonably be conceived to rely upon the idea of self-organization are topics beyond the scope of this paper, it is noteworthy that Kant says that the synthetic unity of perception and moral judgment take place in roughly the same way.[13]

[12] There is no pathway from the receptors or from the brain that can force the formation of such patterns. So these patterns must be the result of interactions between the neurons that comprise the bulb (Freeman 1999, p.71).
[13] Moral judgment, declares Kant, takes place "roughly like the way in which concepts of the understanding, which by themselves signify nothing but lawful form in general, are added to intuitions of the world of sense… (1785, GMS 4:454). A more substantial argument can be made for this assertion. But I defer its treatment to another essay.

Having established that olfaction depends crucially upon the formation of self-organizing spatial patterns in the olfactory bulb, Freeman argues for the pervasiveness of this type of organization in other sensory systems, in the multi-sensory synthetic unity of those systems in experience, and in intentionality.[14] My suggestion is that Freeman's findings lend plausibility to the idea that moral judgment is a function of self-organization in the brain. But Kant's moral theory implies just such an idea. Consequently, Freeman's research lends plausibility to Kant's moral theory.[15] I take the first premise of this argument to be reasonably self-evident. I turn now to defend the second premise.

5. Kant's Model of Moral Judgment

One of Kant's favourite illustrations of moral judgment concerns someone who is holding the property of another in trust (a deposit) the owner of which has died. Suppose, suggests Kant, that the owner's heirs do not know and cannot ever know about the deposit. Present this case even to child of eight or nine and add that, through no fault of his own the trustees fortunes are at their lowest ebb and that his wife and children are sad and suffering from the lack of money. Kant continues:

> And further that the man is kind and charitable, while those heirs are rich, loveless, extremely extravagant spendthrifts, so that this addition to their wealth might as well be thrown into the sea. And then ask whether under these circumstances it might be deemed permissible to convert the deposit to one's own use. Without doubt, anyone will answer "No!" – and in lieu of grounds he can merely say: "It is

[14] Freeman (1999) observes that, "the olfactory, visual, auditory and somatosensory systems, despite having substantial differences with respect to sensation, must have essentially the same mechanisms of perception, because the messages from all these different sensory systems are combined at some level in the brain to form unified multisensory perceptions." (p. 92). Freeman's (1999) discussion of the formation of global AM patterns occurs on pp. 149-155. Herman Haken (2010) has also recently provided support for the role of self-organization in intentionality.

[15] Freeman takes the philosophy of Thomas Aquinas to conform best to his experimental findings (Freeman 2008, p.232). I do not seek to dispute this claim, but rather to add to the list of compatible philosophical systems. I wish also to contribute to the interdisciplinary project of "adjudicating conflicting claims of neuroscientists working on opposing sides of the cleavage: linear-passive versus nonlinear-active perception" (p. 211).

> wrong!" - i.e., it conflicts with duty. Nothing is clearer than that (Kant 1793, TP 8:286-287).

In this example Kant makes no reference to any process that must precede the judgment of the wrongness of the proposed end of keeping the deposit. Moreover, he makes it clear that the individual may be unable to discern grounds for it. Nevertheless, Kant takes this determination to be one made by pure practical reason.

As a first step toward explaining how Kant envisions the process of moral judgment, consider his employment of the same example five years earlier in the *Critique of Practical Reason*. There Kant begins by affirming, as he does in several places, that determining what ought to be done for a given case and what ought not to be done,[16]

> can be distinguished without instruction by the commonest understanding. I have, for example, made it my maxim to increase my assets by every safe means [i.e., I do this as a rule]. Now I have a *deposit* in my hand, the owner of which is deceased and has left no record of it. Naturally, this is a case for my maxim [i.e., my maxim to increase my assets by every safe means]. Now I want only to know whether the maxim can also hold as a universal law (Kant 1788, KpV 5:27).

In this example, to keep the deposit appears to be a safe means to increase my assets. What is there to prohibit me from consistently following through on this maxim and keeping the deposit? In order to distinguish whether the proposed purpose of keeping the deposit is morally permissible I need only to ask whether I can both will my maxim (to increase my assets by every safe means) and, through it, the maxim I that must employ if I am to actualize the purpose of keeping the deposit, but stated in the form of a law that would apply to anyone, i.e., "such a law as this: 'that everyone may deny [the existence of] a deposit which no one can prove to him to have been made." Thus, I need to consider whether I can will that I increase my assets by every safe means, even if this requires that everyone be permitted to lie about the existence of deposits that no one can prove to exist.

[16] Kant is literally concerned with "What form in a maxim is fitting for universal legislation, and what form is not" (Kant 1788, KpV 5:27).

According to Kant, as a result of this exercise, "I immediately become aware that such a principle, as law, would annihilate itself, because it would bring about that there would be no [such] deposit[s] at all" (Kant 1788, KpV 5:27). Evidently, Kant concludes, "my will is subject to a practical law" that requires all my maxims to be capable of functioning as universal laws of the will.[17] This laws states: *act as if the maxim of your action were to become by your will a* **universal law of nature**" (Kant 1785, GMS 4:421).[18]

A law, Kant points out, is "a rule of necessary existence" (Kant 1781/1787, A228/B280). The law of gravity, for instance, is a rule of necessary existence for the solar system. We commonly say that the law of gravity governs the motions of the parts of that system. Yet, strictly speaking, this law is merely a formal description of the rule in accordance with which the parts of that system (must) continuously produce each other with respect to their relative positions; that is, it is a necessary rule in accordance with which the solar system self-organizes. The past, current and future forms of the solar system are all manifestations of this law, which, of course, has nothing to do with special properties of the solar system.

Kant conceives the will along similar lines. He considers the will as a system whose elements produce both each other and, at the same time, a form by means of which those elements appear to be constrained. The elements of the will are its purposes. Since a purpose is, in essence, an idea (e.g., to keep a deposit), any desired purpose (real or imagined) invokes a maxim (a rule) in accordance with which that idea can be made real (e.g., If I do not tell the

[17] As he states elsewhere in the *Critique of Practical Reason*, "through reason we are conscious of a law to which all our maxims are subject, as if through our will an order of nature must at the same time arise" (Kant 1788, KpV 5:44).
[18] emphasis Kant

owners of the deposit then I will be able to keep it.).[19] All maxims necessitate a form of the will, considered as system.[20]

However, the degrees of freedom of that system are limited by the law in accordance with which it self-organizes. As Kant (1788) writes; "(because the concept of causality always contains reference to a law that determines the existence of the manifold [elements] in relation to one another) [the will] has to indicate, as practical reason, *only a law* of objects [of the will]" (KpV 5:89). If the system cannot adopt a form (or gel) in accordance with the dictates of the proposed maxim, then both the maxim and the desired purpose to be attained through it are rejected by the system. Thus, in testing any given maxim, "the mere form of a law, which restricts the matter [i.e., its elements or objects], must at the same time be a basis for adding [or not adding] this matter to the will…" (Kant 1788, KpV 5:34). Moral constraint is, thus, a limit on the degrees of freedom permitted to the will by the law in accordance with which the system self-organizes.[21]

Thinking of the source of moral constraint in this manner leads Kant to declare that in moral judgment all maxims are subject to "a progression…as through the categories of the *unity* of the will's form (its universality), of the *multiplicity* of its matter (its objects – that is, its ends), and of the *allness* or totality of its system of these" (Kant 1785, GMS, 4:436).[22] A plausible interpretation of this passage takes Kant to consider moral constraint to be an emergent property

[19] In the section Kant's *Critique of Judgment* titled "On Purposiveness in General" Kant argues that a purpose, taken by itself, "is the object of a concept insofar as we regard this concept as the object's cause (the real possibility); and the causality that the *concept* has with regard to its *object* is purposiveness (*forma finalis*)" (Kant 1790, KU 5:220).
[20] There is a recent shift in the literature from understanding maxims simply as rules to understanding them as general forms of volition that have the capacity to unify and give direction to a diverse multiplicity of "intentions and actions" (Otfried Höffe, *Ethik und Politik. Grundmodelle und Probleme der praktischen Philosophie*. Frankfurt am Main: Suhrkamp, 1979, in Munzel 1999, pp. 68-9 n.86).
[21] See Bishop (2008) for an interesting discussion of this kind of constraint in Bénard convection. Bishop, for example, writes that "the emergence of the self-regulating large-scale pattern is simultaneous with the modifications of/constrains on the accessible states of motion. … Because of the long-range correlations [or rigidity], an individual fluid element can only execute motions allowed to it by *all other fluid elements* (p. 243).
[22] The three aspects of the "progression" correspond to the three main formulations of the categorical imperative and to three ways of thinking about how the will behaves when a possible purpose is under consideration.

of the total system of purposes as it reacts to a desire to add a new purpose to that system. Since any given purpose only gains purchase on the system by means of a maxim there is a sense in which moral judgment requires that "[practical] reason [as the will] accordingly refers every maxim of the will as giving universal law to every other will and also to every action toward oneself" (Kant 1785, GMS 4:434).[23] Given the complete interdependence of everything in this system Kant concludes that the "determination of the will... [in a moral judgment] ties the concept of causality to conditions that are entirely different from those that amount to natural connection [i.e., to mechanism]" (Kant 1787, KpV 5:69).[24]

According to Kant, it is precisely the inability of previous philosophers to think of the will in this manner that accounts for their failure to discover the supreme principle of morality. Their problem, quite simply, is that they always conceive the supreme good as a supreme purpose, which must then be attained. Kant (1787) writes:

> Now, whether they posited this object of pleasure – which was to yield the supreme concept of the good – in happiness, in perfection, in moral feeling or in the will of God,... they could call their object – as direct determining basis of the will – good or evil only according to its direct relation to feeling, which is always empirical (KpV 5:65).

In short, for any hypothesized supreme good one can always ask: "How do I feel about it?" But the determination of the will takes place "with the aim" of subjecting "a priori the manifold of *desires* to the unity of consciousnesss of a practical reason commanding in the moral law or [i.e.] of a pure will." (Kant 1787, KpV 5:65)[25] In other words, it is in the nature of the will to subject

[23] "Every other will" indicates all possible states of a coherent will. The act of referring every maxim to every other possible maxim can be conceived in terms of a history of attractor landscapes of a system, where all past states of the system (all previous purposes, real or imagined) have determined its current form.

[24] In the *Critique of Judgment*, Kant notes that "when we use the word *cause* with regard to the supersensible, we mean only the ground (*Grund*) that determines *natural things* to exercise their causality to produce an effect in conformity with the natural laws proper to that causality, yet at the same time in accordance with the formal principles of the laws of reason" (Kant 1793, KU 196).

[25] When one has a desire, it is a desire for something. Consequently, Kant means the same thing by "the manifold of desires" as by "the manifold of objects of the will;" that is, of purposes.

every desired purpose, including the supreme purposes of happiness, perfection, moral feeling, and the will of God, to the rational constraint implicit in the total system of purposes.[26] As Kant observes in *the Groundwork of Metaphysics of Morals*, the previous attempts of philosophers to find the supreme principle of morality had to fail because "it never occurred to them that man is subject only to laws which are made by himself and yet are universal, and that he is bound only to act in conformity with a will which has as its natural purpose lawful self-determination" (Kant 1785, GMS, 4:432).[27] Thus, Kant succeeds where others fail because he thinks of the will as a special kind of a self-organizing system (i.e., one that acts in accordance with principles), whose (as if) purpose in nature is to produce a will that is good (see Kant 1790/1793, KU 5:373-376).

6. Conclusion

Although the debate on moral foundations continues to be framed in terms of the dichotomy of feeling or reason, that situation may be changing. First, a recent comprehensive re-evaluation of the relevant neuroimaging data argues that it does not support that dichotomy. Instead the data supports the existence of a single-unified system of moral evaluation that is "instantiated by a single, complex brain network" that, in planning for action, engages "an alternative, non-actual environment" (see Klein 2011, pp.152 and155-7). Second, this essay provides reason to doubt the support that reaction time data provides to that dichotomy because it

[26] "Rational nature is distinguished from the rest of nature by this, that it sets itself an end. This end would be the matter of every good will. But since, in the idea of a will absolutely good without any limiting condition (attainment of this or that end) abstraction must be made altogether from every end to be *effected* [i.e., actualized] (this would make every will only relatively good), the end here must be thought not as an end to be effected but as an *independently existing* (*willkürlichen*) end, and hence thought of only negatively, that is, as that which must never be acted against and which must therefore in every volition be estimated never merely as a means but always at the same time as an end" (Kant 1785, GMS 4:437).

[27] Kant writes: "*dem Naturzwecke nach aber allgemein gesetzgebenden Willen gemäß zu handeln*" (lit. according to the natural purpose, the will is determined by the general legislating will). In this passage the phrase *dem Naturzwecke nach* is commonly translated as "purpose of nature" (*Zweck der Natur*), for which translators then occassionally seek to provide *ad hoc* justification (see, for example, H.J. Paton's translation of the Groundwork of Metaphysics of Morals, p. 100 n.).

introduces a third way to account for that data, i.e., for the rapidity and automaticity of moral judgment. This is significant because the main argument from this data in support of emotion as the cause of moral judgment is simply a disjunctive syllogism based on that dichotomy. Consequently, to introduce a third option is to demonstrate that it is false dichotomy and, hence, to undermine the soundness of the argument.

Before concluding this essay some speculative comments are in order concerning the neural tenability of Kant's theory of moral judgment. This theory, as I have explicated it, hypothesizes a self-organizing network of purposes that generates moral constraint in response to a desired additional purpose. Such a theory has several virtues, one of which is that accounts for the possibility of the emergence of objective moral constraint from merely subjective conditions. Just as Bénard cells will emerge in all sorts of liquids, this theory renders possible a universal moral basis in individuals with extremely variable sets of individual purposes .

It is, however, implausible that the desire to attain a particular purpose can literally invoke the interaction of all past or all possible purposes. To help make sense of this aspect of the theory I appeal for ideas to Freeman's discussion of the attractor landscape jostling that occurs in the olfactory bulb upon exposure to a new and significant odorant. Freeman argues that,

> each amplitude modulation pattern depends on the history of exposure not merely to one odorant, but to every odorant…there is a single large attractor for the olfactory system, which has multiple wings that form an attractor landscape. This attractor landscape contains all the learned states as wings of the olfactory attractor. When a new class is learned, the synaptic modifications in the neuropil jostle the existing basins of the packed landscape…This is known as attractor crowding. No basin is independent of the others (1999, p. 107-8).

It may be possible to extend some aspects of this type of non-linear systems account of the dynamics of the olfactory bulb to the neural instantiation of a system of purposes (see also Carver and Scheier 2002). In this case, the single large attractor may correspond to the law in

accordance with which the system self-organizes. The multiple wings may correspond to maxims invoked in consideration of past desired purposes (real or imagined). It may also be possible to integrate aspects of Freeman's discussion of intentionality.

Finally, it may be objected that the theory of moral judgment that I have presented requires that every purpose that one sets be subject to moral evaluation. Since it is quite clear that this is not the case – that the majority of purposes are set and executed on auto-pilot and without the slightest hesitation – the theory is implausible. The problem with such an objection is that its first premise is false. Just as most smells pass us by relatively unnoticed, so most purposes that we set and execute occur at a very low level of attention. Just as the invocation of the determinative powers of the olfactory bulb depend upon the odorant being significant, so also the invocation of the determinative powers of the moral system depends upon the proposed purpose being significant. In both cases, desire sets the neural gain sufficiently high to engage the services of the system. Needless to say, the proposed purpose of keeping the deposit in Kant's example is amply significant, since its actualization has the capacity to alleviate considerable personal suffering. There is, by contrast, no reason to think that a purpose to pour myself a cup of coffee will trigger the moral system, unless the coffee is, for example, not mine and I "need" one quite badly. In this case, my mother's slap on the hand in a relevantly similar situation of thirty years ago may generate an emotional reaction, which could be mistaken for a moral judgment. At the same time, this emotion may prompt *reflection*, which may or may not involve deliberation, by means of which I judge for myself whether I ought to feel badly about such an act (see Kant 1790/1793, FI 211'). If Kant's moral theory is correct then moral judgment does not even require conscious isolation of one's maxim, although that is the best way to proceed, since it provides a method of testing the soundness of the underlying process.

References

Bargh J. A. and Ferguson M. L. (2000): Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin* 126, 925-945.

Bishop R.C. (2008): Downward causation in fluid convection, *Synthese* 160, 229-248.

Carver, C. S., & Scheier, M. F. (2002): Control processes and self-organization as complementary principles underlying behavior. *Personality and Social Psychology Review*, 6, 304 –315.

Cushman F. A., Young, L. and Hauser M. D. (2006): The role of conscious reasoning and intuitions in moral judgment: testing three principles of harm. *Psychological Science* 17(12), 1082-1089.

Damasio A. R. (1994): *Descartes' Error: Emotion, Reason and the Human Brain*, NY, Grosset/Putnam.

Freeman W.J. (1999): *How Brains Make Up Their Minds*, Weidenfeld and Nicolson, London.

Freeman W.J. (2003): A neurobiological theory of meaning in perception. Part 1. Information and meaning in nonconvergent and nonlocal brain dynamics. *AM. Int. J. Bifurc. Chaos* 13, 2493-2511.

Freeman W.J. (2008): Nonlinear Brain Dynamics and Intention According to Aquinas. *Mind and Matter* 6(2), 207-234.

Greene J. D., Nystrom, L. E., Engell, A. D., Darley, J. M. and Cohen, J. D. (2004): The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44, 389-400.

Greene J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M. and Cohen, J. D. (2001): An fMRI investigation of emotional engagement in moral judgment. *Science* 293, 2105-2108.

Greene J. D. (2007): 'The Secret Joke of Kant's Soul'. In W. Sinnott-Armstrong, ed., *Moral Psychology, Vol. 3: The Neuroscience of Morality: Emotion, Disease, and Development*, Cambridge, MA: MIT Press, 35–79.

Haken H. and Tschacher W. (2010): A Theoretical Model of Intentionality with an Application to Neural Dynamics. *Mind and Matter* 8(1), 7-18.

Haidt, J. (2001): The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review* 108, 814-834.

Haidt J., and Joseph C. (2004): Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues. *Daedalus*, Special issue on human nature, 55-66.

Haidt J. and Bjorklund F. (2007): Social intuitionists answer six questions about moral psychology. In W. Sinnott-Armstrong (ed.), *Moral Psychology, Vol. 2: The Cognitive Science of Morality: Intuition and Diversity*, Cambridge, MA, MIT Press, 181-217.

Höffe O. (1979): Ethik und Politik. Grundmodelle und Probleme praktischer Philosophie, Suhrkamp, Frankfurt.

Hume D, (1739): *A Treatise of Human Nature*, ed. by Selby-Bigge, Oxford University Press 1888.

Hume D. (1751): *An Enquiry Concerning Principles of Morals*, ed. by J.B. Schneewind, Hackett, Indianapolis 1983.

Kant I. (1781/1787): *Critique of Pure Reason*, transl. by W.S. Pluhar, Hackett, Indianapolis 1996.

Kant I. (1785): *Groundwork of the Metaphysics of Morals*, transl. by M. Gregor, Cambridge University Press, NY 1997.

Kant I. (1785): *Groundwork of Metaphysics of Morals*, transl. by H.J. Paton, Harper Perennial, 1964.

Kant I. (1788): *Critique of Practical Reason*, transl. by W.S. Pluhar, Hackett, Indianapolis 2002.

Kant I. (1790/1793): *Critique of Judgment*, transl. by W.S. Pluhar, Hackett, Indianapolis 1987.

Kant I. (1793): *On the old saw: that may be right in theory but it won't work in practice*, transl. by E.B. Ashton, University of Pennsylvania Press, Philadelphia 1974.

Kant I. (1797): *The Metaphysics of Morals*, transl. by M. Gregor, Cambridge University Press, NY 1996.

Klein C. (2011): The dual-track theory of moral decision making: A critique of the neuroimaging evidence. *Neuroethics* 4, 143-162.

Leben D. (2011): Cognitive neuroscience and moral decision-making: Guide or set aside? *Neuroethics* 4, 163–174.

Lapsley D.K. and Hill P.L. (2008): On dual processing and heuristic approaches to moral cognition. *Journal of Moral Education* 37(3), 313-332.

Munzel G. F. (1999): *Kant's Conception of Moral Character: The "Critical" Link of Morality, Anthropology and Reflective Judgment*, The University of Chicago Press, Chicago.

Murphy F.C., Wilde G., Ogden N., Barnard P.J. and Calder, A.J., (2009): Assessing the automaticity of moral processing: Efficient coding of moral information during narrative comprehension. *The Quarterly Journal of Experimental Psychology* 62 (1), 41-49.

Pessoa L. (2008): On the relationship between emotion and cognition. *Nature Reviews Neuroscience* 9(2), 148–158.

Quian Quiroga R. and Panzeri S. (2009): Extracting information from neuronal populations: Theory and decoding approaches. *Nature Reviews Neuroscience 10*, 173-185.

Rizzolatti G. And Craighero L. (2004): The mirror-neuron system. *Annual Review of Neuroscience* 27, 169-192.

Schnall S., Haidt J., Clore G.L., Jordan A.H. (2008): Disgust as Embodied Moral Judgment, *Pers Soc Psychol Bull* 34(8), 1096–1109.

Sober E. and Wilson D.S. (1998): *Unto Others: The Evolution and Psychology of Unselfish Behavior*, Harvard University Press, Cambridge, MA.