Explaining Freedom in Thought and Action

Patricia Kitcher
Columbia University

1. Kant's Goals

Kant's central projects in epistemology and in ethics manifest his distinctive approach to philosophy. In epistemology, he aims to reveal the necessary conditions for the possibility of empirical cognition, in ethics, the necessary conditions for the possibility of moral action. Although there are important differences between the cases, I will emphasize the similarities in his approaches to epistemology and to ethics for two reasons. One is that, as is frequently noted, he seemed to believe that his investigations into 'freedom' in thought could in some way be carried over to the sphere of action. My second reason is that taking his investigations of the necessary conditions for the possibility of empirical cognition as a model for his analysis of the necessary conditions for moral action makes it easier to avoid a blind alley into which several texts entice the unforewarned reader.

The blind alley is this: Kant is trying to analyze the necessary conditions for the possibility of moral or free action. Since the target is action, as opposed to mere bodily movement, the question of the possibility of free action can quickly be reduced to that of the possibility of free choice of action. If that is Kant's *analysandum*, however, then the *analysans* cannot include the 'free choice' of a *Gesinnung* by a noumenal agent (Wood,1984, 89ff.) or by an immature agent (Allison, 1990, 136ff.) or the self-conscious, autonomous decision by "something over and above all of your desires, something that is you" (Korsgaard, 1989, 323)[1]. The project makes sense only if the possibility of free choice requires analysis and explanation; in that case, however, the results cannot include an unexplicated notion of 'free choice.'

Another way to see the problem is in the language of philosophy of psychology. The proposed explications would be 'homuncular.' They explain how a human being is able to do something—make free choices—by appealing to an inner *homunculus* who makes free choices. By contrast, in proper functional analysis, the higher level capacity

is explained in terms of sub-capacities that carry out *different* tasks, the sum total of which are constitutive of doing the whole task. Keeping Kant's analyses of cognition in mind is helpful, because there is no temptation to think that after sensory representations have been brought in according to the forms of intuition and consciously combined according to rules associated with concepts, thereby producing a representation—that representation is then *cognized* by a noumenal I.

Beyond laying out Kant's analysis of the necessary conditions for the possibility of morality, I briefly sketch a scientific account that would explain how humans can meet those conditions. This account provides a basis for a metaphysics of moral psychology that differs from that of phenomena and noumena. Kant introduces his metaphysics to reconcile two claims: The complete state of the world at each moment in time determines the subsequent state; there is more than one possibility for the state of the world at the next moment in time. These doctrines seem irreconcilable to me, however many worlds or aspects we talk about,[2] so the scientifically inspired metaphysics that I propose to support his moral psychology is not intended to handle this issue. It is addressed instead to a second reason that Kant has for believing that the necessary conditions for the possibility of morality cannot be fulfilled in a world that can be explained by natural science. I argue that this second justification for noumenalism in ethics can now be rejected, thus permitting us to adopt a more plausible metaphysics that can support many elements of his moral psychology.

2. *Rational Desire*

Despite his commitment to systematic presentation, Kant discusses willing in both the *Critique of Pure Reason* and the *Groundwork of the Metaphysics of Morals* before indicating the basic power or faculty required for it. As he finally explains in the Preface of the *Critique of Practical Reason*, that basic faculty is desire, which is characterized in a note as

> das Vermögen desselben [of the living being], durch seine Vorstellungen Ursache von der Wirklichkeit der Gegenstände dieser Vorstellungen zu sein. (5.9a)

He excuses his failure to be explicit about this basic faculty in the *Groundwork* on the grounds that it was reasonable for him to presuppose standard psychological

classifications. On the other hand, he also explains that he disagrees with a key piece of the standard account. According to that account, the feeling of pleasure is assumed always to be the determining basis of the faculty of desire and that assumption cannot be made when considering ethics, since it would make morality impossible (5.9a).

With this *post hoc* background available, we may turn to the *Groundwork's* familiar introduction of the will. Animals as well as humans have a faculty of desire that can produce objects through representations. But they do it differently:

> Nur ein vernünftiges Wesen hat das Vermögen, nach der Vorstellung der Gesetze, d. i. nach Principien, zu handeln, oder einen Willen. Da zur Ableitung der Handlungen von Gesetzen Vernunft erfordert wird, so ist der Wille nichts anders als praktische Vernunft. (4.412)

Where the representations of animals lead to actions through the law of association, in the case of rational beings, it is the representation of a law *per se*—a principle understood as such—that leads to the production of action. In rational desiring that production takes a special form: The action is derived from the principle or borrowing an expression from Sebastian Rödl, the thought governing bodily movement is derived from a principle (2007, 17).

Kant's discussions of willing in the *First Critique* contrast the desiring of animals with that of humans in a different way, in terms of freedom:

> Eine Willkür nämlich ist bloß thierisch (*arbitrium brutum*), die nicht anders als durch sinnliche Antriebe, ... bestimmt werden kann. Bewegursachen, welche nur von der Vernunft vorgestellt werden, bestimmt werden kann, heißt die freie Willkür (*arbitrium liberum*) ... Die praktische Freiheit kann durch Erfahrung bewiesen werden. Denn nicht bloß das, was reizt, d. i. die Sinne unmittelbar afficirt, bestimmt die menschliche Willkür, sondern wir haben ein Vermögen, durch Vorstellungen von dem, was selbst auf entferntere Art nützlich oder schädlich ist, die Eindrücke auf unser sinnliches Begehrungsvermögen zu überwinden … (A802/B830).

Whereas the faculty of desire in animals is stimulus bound, it is evident through experience that humans can overcome current desires by considering their long-term interests.

This passage seems inconsistent. It opens with a contrast between creatures who can only be determined through impulses and (presumably) those who can also be determined by something else. Yet, experience only shows that humans are not bound

by immediate impulses, not that they can act without any sensory impulse at all. Perhaps one purpose of the later discussion of the faculty of desire is to remove this confusion: What is crucial for the possibility of morality is that the determining ground of the faculty that produces action is not always pleasure or something sensible.

That this is the crucial point for the possibility of moral action is clear from the *Groundwork*. To have moral worth an action must be done from the motive of duty and not from any inclination. It is constitutive of moral goodness that

> nichts anders als die Vorstellung des Gesetzes an sich selbst, die freilich nur im vernünftigen Wesen stattfindet, so fern sie … der Bestimmungsgrund des Willens ist … (4. 401)

That is, moral action is possible only if the representation of the moral law is the determining ground of the will. When Kant turns to the question of the *Triebfeder* or spring of moral action in the *Critique of Practical Reason*, this thesis is not altered, but reinforced. The only *Triebfeder* for moral action that would not undermine its status is the moral law itself, even though it is an insoluble problem for human reason to understand how the moral law can directly determine the will:

> wie ein Gesetz für sich und unmittelbar Bestimmungsgrund des Willens sein könne (welches doch das Wesentliche aller Moralität ist), das ist ein für die menschliche Vernunft unauflösliches Problem und mit dem einerlei: wie ein freier Wille möglich sei. (5. 72)[3]

If it is impossible for humans to know how the representation of the moral law can be a determining ground of the faculty of rational desire, then the sort of causation involved must not fit the parameters that are suitable for human cognition. A Reflection indicates the character of the mismatch:

> Bewegungen können also … durch nichts, was nicht selbst ~~bewegt~~ vorher bewegt war, anfangen (R5997, 18. 421)[4]

The problem with understanding how rational principles could be the determining ground of the will is that they cannot be understood in terms of motions that are communicated from one object to another. And as Kant makes clear in another Reflection, that is what would be necessary to comprehend the possibility of free or moral action:

> In den freyen Handlungen fließt die Vernunft nicht blos als ein begreifendes, sondern wirkendes und treibendes *principium* ein. Wie sie nicht blos vernünftle und urtheile, sondern die Stelle

> einer Naturursache vertrete, sehen wir nicht ein, viel weniger, wie sie durch Antriebe selbst zum handeln werde. (R5612, 18. 253)[5]

Precisely what is needed for the possibility of moral action—that reason be an active or efficient cause of movement through its moral law—cannot be understood, because principles of reason cannot be understood in terms of transfer of motion.

I appeal to these Reflections to provide evidence that Kant's conviction that there must be two worlds or two aspects depends in part on an assumption about science. Rational causes cannot be brought within the bounds of scientific explanation, because science revolves around the laws of motion.[6]

3. The Proofs of Freedom and the Moral Law

Given Kant's conviction that it is impossible to explain *how* the moral law could determine a rational faculty of desire, his demonstrations that humans are free and that morality is not a chimera are limited to showing *that* it can do so. Although their primary purpose is to establish that humans have the capacities required for morality, these demonstrations also provide further specifications of those capacities. They are presented in the third section of the *Groundwork* and in the so-called 'fact of reason' passages in the *Second Critique*.[7]

In *Groundwork 3*, Kant defines 'freedom' of the will in terms of independence from determination by alien causes (4.446). He then appeals to freedom in thought to try to demonstrate a possible freedom in acting. It is impossible

> sich … eine Vernunft denken, die mit ihrem eigenen Bewußtsein in Ansehung ihrer Urtheile anderwärts her eine Lenkung empfinge, denn alsdann würde das Subject nicht seiner Vernunft, sondern einem Antriebe die Bestimmung der Urtheilskraft zuschreiben. (4.448)

Suppose that Kant is right that *if* a subject were conscious of receiving direction from an alien source, then she would attribute the determination of her judgment to an impulse and not to reason. How can this hypothetical claim establish that human judging is free from alien impulses?

I assume that Kant is drawing on his discussion of rational cognition in the *Critique of Pure Reason*. There he argues that judging is possible only if the subject has a faculty of understanding that consciously combines representations in further

representations.[8] Given that account, anyone who makes a judgment must be conscious of the basis on which she produces it. It follows that anyone who makes judgments on the basis of principles and evidence would know that she does. She would also know that she produces judgments independently of alien influences. So anyone who enjoys freedom in thought would know that she does.

Although there are many hypotheses about why Kant thought this argument failed to establish the reality of morality, two problems seem fairly obvious. First, even if a subject knows that she can produce *judgments* through her intellectual grasp of principles independently of sensible impulses that does not give her any cognition that she can also produce judgments capable of guiding action solely on the basis of principles. Second, even if humans can act on principle that would not show that the human faculty of rational desire can be determined by a principle with the content of the moral law.

One reason I highlight these objections to the argument of *Groundwork* 3 is that the discussion in the fact of reason passages seems designed to block them. The line of argument in the *Second Critique* is surprisingly direct. As noted, the Preface raises the issue of whether only pleasure can be the determining ground of the will—a result that would show morality to be impossible. The Introduction returns to this issue noting immediately that

> Hier ist also die erste Frage: ob reine Vernunft zur Bestimmung des Willens für sich allein zulange, oder ob sie nur als empirisch=bedingte ein Bestimmungsgrund derselben sein könne. (5. 15)

That is, the first or central question is (again) whether not only pleasure, but also reason can determine the faculty of desire. That is the question to which the fact of reason discussion is supposed to provide a positive answer.

The 'fact of reason' texts have been the subject of constant criticism. I will not offer much defense here,[9] and will only use them to lay out more of Kant's moral psychology. His claim is that when an agent considers a course of action, she is always conscious of the moral law. Given his remarks in the *Groundwork* (4.402) and the *Second Critique (*5.69), he does mean that she is conscious of one of the precisely crafted formulae of the categorical imperative, but rather of something such as 'what if

everybody did that?' Further, she is conscious of it as determining her willing independent of sensibility:

> [Es] ist … das moralische Gesetz, dessen wir uns unmittelbar bewußt werden (so bald wir uns Maximen des Willens entwerfen), welches sich uns zuerst darbietet und, indem die Vernunft jenes als einen durch keine sinnliche Bedingungen zu überwiegenden, ja davon gänzlich unabhängigen Bestimmungsgrund darstellt … (5. 29-30)

Since what Kant needs to show is that reason can determine the will on its own, the bald assertion that humans are conscious of their reasoning as doing so can prompt memories of Russell's warning about the advantages of theft over honest toil.

Again, however, we can get some help in following Kant's reasoning by looking back to the *First Critique*. There he argued that since humans have no access to thinking through either outer or inner sense, the only way that they can cognize it is by engaging in it.[10] (As we have just seen, it is by engaging in thinking that cognizers know that they enjoy freedom of thought.) If he is right that this is the only way that we can cognize higher mental processes, then there would be no way to prove that the moral law moves the willing other than through subjects' experience of practical deliberation. Further, Marcus Willaschek has argued that the cases of lust and false testimony that follow this discussion should be understood as thought experiments that the reader is invited to perform in order to confirm through his own deliberating that his will can be moved by moral considerations. The false testimony case is very stark: All of the advantages of pleasure are on the side of providing the Prince with false testimony, yet Kant is confident that his readers will be conscious of being repelled by the thought of doing so (5.30).

We now have a fairly full analysis of the necessary conditions for the possibility of moral action: Morality is possible only if humans have a faculty of desire that is not bound to react to immediate stimuli, a faculty of reason that can derive particular actions from general principles independently of sensible impulses, a faculty of desire that is positively free in that it can moved by consciousness of the moral law and negatively free, because if the faculty of desire is moved by the moral law, then it is moved independently of any sensible impulses. Further, Kant has offered *Gedanken* experimental demonstrations to show that humans have the required faculties.

Why aren't we done? I think that we are basically done, although I will consider extra conditions for immoral actions in the next section. The problem is that Kant's moral psychology can be confusing, because of issues that he considers later, specifically his introduction of the notion of *Gesinnung* in the *Religion* book and his clarification of the relation between *Wille* and *Willkür*.[11]

4. *Why Free Action Should Not be Analyzed in terms of a Noumenal Self's Free Choice of a Gesinnung*

Although it plays a prominent role in *Religion within the Bounds of Mere Reason*, the notion of a *Gesinnung* (or fundamental attitude) is not given any serious treatment in either the *Critique of Practical Reason* or the *Metaphysics of Morals*.[12] Still, one occurrence of *Gesinnung* in the *Second Critique* may give the source of a general issue in moral psychology that Kant tries to straighten out in the *Religion* book *via* the introduction of this expression. The end of the discussion of the highest good refers to the battle that

> die moralische Gesinnung mit den Neigungen zu führen hat (5.147).

As Gerold Prauss (1983, 92) and others[13] have noted, the *Second Critique* errs in suggesting that there is a conflict between the moral law and inclinations, since there is nothing morally wrong with inclinations.

One advance of the *Religion* book is to fix this mistake by providing an analysis of how morally bad or evil action is possible.[14] Kant's account draws out the implications of the obvious requirement that no one can act immorally who does not know how to act correctly. In that case, the necessary conditions for the possibility of morally correct action must be presupposed in explaining the possibility of morally bad action. He introduces the notion of *Gesinnung* in part to explain how the conditions for the possibility of immoral actions rely on the conditions required for morally good actions.

How does a person act badly? Like the good agent, the bad one must be conscious of the moral law. Further, as is implied in the fact of reason discussion and made explicit in the *Metaphysics of Morals* (6.399-400), his consciousness of the moral law is not passive, but must have effects. He must be conscious of feeling attracted to or repelled by an action through considering whether everyone could do what he proposes

to do. Finally he must be conscious of thus changing his volitional structure through considering the moral law independently of any inclination. If he were an angel, unaffected by sensible desires, then the alteration in his willing whereby he is repelled by the action would lead straight to omitting the action.[15] So, immoral action is possible only if there is some counterweight that can oppose the motive produced by consciousness of the moral law.

Since we may assume that humans are universally motivated by considerations of present and future happiness,[16] we know that there is such a counterweight: rational self-interest. However, insofar as a subject is capable of moral action (and so of immoral action), he must be conscious of the moral law as moving his will independently of inclination, and so recognize that the demands of morality are independent of inclination whereas those of rational self-interest are conditional on it.

Immoral action is possible because although when an agent engages in moral deliberation, she is conscious in moving her willing through her consciousness of the moral law, she can also turn away from that consciousness. Kant makes this point after introducing the formula of universal law in *Groundwork 1*. Because humans also have plans to further their happiness, they concoct rationalizations to convince themselves that the strict laws of duty are not valid (4.405). In the Preface to the *Second Critique*, he raises the possibility that people will try to rationalize away the power of the moral law to move them to act.[17] The language of the *Religion* book is slightly different—immorality is possible only because people pervert what they know to be the case by reversing the priority of the moral law over self-interest—but the point is the same. Moral agents know that the demands of the moral law are independent of inclination, but they treat them as if they were dependent on their desires. That is, immoral action is possible only through self-deception.[18]

The notion of a *Gesinnung* that prioritizes either the moral law or self-interest is helpful, because it enables Kant to give a completely general account of immorality. But the *Religion* book also has several discussions that tempt us to read him as maintaining that an act can be free and imputable, only because a noumenal or timeless self freely chooses a *Gesinnung*. One passage is where he introduces the notion of a *Gesinnung*:

> Die Gesinnung, d. i. der erste subjective Grund der Annehmung der Maximen, kann nur eine einzige sein und geht allgemein auf den ganzen Gebrauch der Freiheit. Sie selbst aber muß auch durch freie Willkür angenommen worden sein, denn sonst könnte sie nicht zugerechnet werden. (6.25)

This text can easily be read as suggesting a two stage process: First an agent chooses a *Gesinnung* freely and then that fundamental attitude directs her free action.

This interpretation cannot be correct, however, if what is meant is that the agent selects her *Gesinnung* by exercising a faculty that enables her to choose 'as she pleases.' As will be even clearer in the next section, Kant maintains that a faculty of desire is free only if it can be determined by reason's moral law (6.213). In that case, the possibility of a free act is not explained by a free choice of a *Gesinnung*; rather, since action is bodily movement guided by a maxim, the necessary conditions for the possibility of a free choice of a maxim or a *Gesinnung* are the same as the necessary conditions for the free choice of action (minus the physical movement), *viz.*, that the faculty of desire not be stimulus bound, that it be positively and negatively free, *etc.* These necessary conditions are neither made redundant, nor added to by the introduction of the notion of a *Gesinnung*. What is new is not that it is a necessary condition for free and moral action that the agent freely chooses her maxim, but the idea that agents' maxims all reflect a constant fundamental attitude to prioritize morality or self-love. This doctrine is a reflection of Kant's rigorism, which I'm not going to explore.

A second passage concerns the relation between a person's *Gesinnung* and her actions.

> dieses [ein Hang[19] zum Bösen] muß aus der Freiheit entspringen … versteht man unter dem Begriffe eines Hanges einen subjectiven Bestimmungsgrund der Willkür, der vor jeder That vorhergeht, mithin selbst noch nicht That ist; da denn in dem Begriffe eines bloßen Hanges zum Bösen ein Widerspruch sein würde, wenn dieser Ausdruck nicht etwa in zweierlei verschiedener Bedeutung, die sich beide doch mit dem Begriffe der Freiheit vereinigen lassen, genommen werden könnte. Es kann aber der Ausdruck von einer That überhaupt sowohl von demjenigen Gebrauch der Freiheit gelten, wodurch die oberste Maxime (dem Gesetze gemäß oder zuwider) in die Willkür aufgenommen, als auch von demjenigen, da die Handlungen selbst (ihrer Materie nach, d. i. die Objecte der Willkür betreffend) jener Maxime gemäß ausgeübt werden. Der Hang zum Bösen ist nun That in der ersten Bedeutung (*peccatum originarium*) und zugleich der formale Grund aller gesetzwidrigen That im zweiten Sinne genommen, welche der Materie nach

> demselben widerstreitet und Laster (*peccatum derivativum*) genannt wird; und die erste Verschuldung bleibt, wenn gleich die zweite (aus Triebfedern, die nicht im Gesetz selber bestehen) vielfältig vermieden würde. Jene ist intelligibele That, bloß durch Vernunft ohne alle Zeitbedingung erkennbar; diese sensibel, empirisch, in der Zeit gegeben (*factum phaenomenon*). (6.31)

The view that a necessary condition for free action is a timeless choice of *Gesinnung* is encouraged by the claim at the end of the passage that the inscrutability of free choice is rooted in its timelessness. Notice, however, that that claim is in tension with earlier claim that the outward act or deed must be 'preceded' by an inner act of adopting a *Gesinnung* for or against the moral law.

As Henry Allison has noted, the obvious way around the tension is to see the priority of choosing a *Gesinnung* as 'logical' (1990, 143-44). That is, the free choice of a *Gesinnung* is a necessary condition for the possibility of moral action: It is part of Kant's analysis of the necessary conditions for moral action that it is action guided by a maxim that is chosen under conditions of positive and negative freedom and a *Gesinnung* is something like a meta-maxim.[20]

I have not tried to deal with the issue of moral rigorism that is in play in the first passage or with Kant's attempt to connect his views to the doctrine of original sin that is at issue in the second. My concern has only been to show that insofar as it draws on or implies claims in Kant's general moral psychology, the notion of a *Gesinnung* can be understood in a way that is consistent with seeing him as offering an informative analysis of how a human being acts morally or, more particularly, immorally—namely through altering his volition through his rational grasp of the moral law and through deceiving himself that what he is conscious of as an unconditional demand is conditional on desires. Kant does not make the error of explaining the possibility of human free action through appealing to the unexplicated free choice of a *Gesinnung* by a self within the self.

5. *Why Free Choice Should Not be Explained through a Free Faculty of Choice*

Kant's introduction of the crucial issue of acting on principles contains a contradiction. After identifying '*Wille*' as 'nichts anders als praktische Vernunft,' he continues:

> Wenn die Vernunft den Willen unausbleiblich bestimmt ...., d.i. [wenn] der Wille ist ein Vermögen, **nur dasjenige** zu wählen, was die Vernunft unabhängig von der Neigung als praktisch nothwendig, d. i. als gut, erkennt. (4.412)

If the *Wille* is identical with practical reason, then how could it *not* choose what reason recognizes as good?

This passage and others have led Kant's contemporaries and many subsequent readers to object that he has ruled out the possibility of immoral action.[21] According to a popular view, he finally solves the problem in the *Metaphysics of Morals* with a distinction between two kinds or aspects of will, *Wille* and *Willkür*.[22]

> Von dem Willen gehen die Gesetze aus; von der Willkür die Maximen. Die letztere ist im Menschen eine freie Willkür; der Wille, der auf nichts Anderes, als bloß auf Gesetz geht, kann weder frei noch unfrei genannt werden, *weil er nicht auf Handlungen*, sondern unmittelbar auf die Gesetzgebung für die Maxime der Handlungen (also die praktische Vernunft selbst) *geht*, daher auch schlechterdings nothwendig und selbst keiner Nöthigung fähig ist. Nur die Willkür also kann frei genannt werden. (6.226)

Lewis White Beck read this and other passages as indicating that *Wille* and *Willkür* should be understood on the model of a single will with a legislative (*Wille*) and an executive (*Willkür*) branch (1960, 180).

We have already seen the problem with this approach in the discussion of *Gesinnung*. The necessary conditions for a free choice of a *Gesinnung* are exactly the necessary conditions for the possibility of a free or moral action (minus bodily movement). Putting the focus on an alleged 'faculty'[23] of choice, *Willkür*, rather than on the object of choice does not change this basic fact. *Willkür*'s choice can be free only if the necessary conditions for free action are met (again, minus bodily movement). In that case, however, we are requiring that a sub-capacity that is supposed to fulfill some necessary conditions for the possibility of moral or free action meet all the conditions required for the original capacity. Or we are helping ourselves to an unanalyzed notion of free choice that undermines the point of the original project of analysis.

In his systematic presentation in the *Metaphysics of Morals,* Kant provides an explicit account of the notion of 'free choice,' an account that is already at work in the *Groundwork* and the *Practical Critique*:

> Die Willkür, die durch reine Vernunft bestimmt werden kann, heißt die freie Willkür (6.213).

It follows that only some types of *Willkür* are free, namely those that can be determined by a *Wille*. A *Willkür per se* is a kind of faculty of desire that can be moved by representations, and so concepts and principles, and that is not compelled by immediate inclination. It can do or refrain as it pleases.

A human faculty of desire capable of producing free action must have further properties. His technical notion of '*Wille*' enters as a further specification of the kind of *Willkür* that humans have,[24] namely one whose action-motivating feelings can be determined by reason's grasp of the moral law.

Immediately after contrasting the unfreedom of *Wille* with the freedom of *Willkür*, Kant warns against the sort of reading that Beck and others[25] have endorsed:

> Die Freiheit der Willkür aber kann nicht durch das Vermögen der Wahl, für oder wider das Gesetz zu handeln, (*libertas indifferentiae*) definirt werden - wie es wohl einige versucht haben … (6.226)

Current commentators believe that the target of this remark is Carl Leonhard Reinhold (Allison, 133, Wuerth, 237, Franks, 270, n.12, cf. Prauss 1983, 83ff.), whom Beck did not reference. Reinhold had tried to improve on Kant's theory by suggesting that free action requires just such a choice:

> [Die Freiheit] …ist mehr als die unwillkührliche von Selbstthätigkeit der practischen Vernunft, durch welche nichts als das bloβe Gesetz gegenben wird … sie ist die willkührliche von practischen Vernunft wesentlich verschiedene Selbsttätitgkeit der Person, durch welche das Gesetz entweder ausgeführt oder übertreten wird (BKP, vol. 2. 297)

Kant objects that Reinhold has failed to provide an acceptable definition of this novel faculty that can execute or contravene the dictates of moral law. The problem is that although humans can know that they can alter their wills independently of sensory influences, merely through being conscious of the moral law, they have no idea how.

To appreciate Kant's point against Reinhold, imagine that there were a theory of how humans change their desiderative structures through contemplating the moral law: They do it by X'ing. Under those circumstances, it would be possible to

give a principled account of a capacity that acted for or against the moral law—because that would be a capacity that enhances or weakens X'ing. Since no such account is possible, we are left with nothing but empirical hypotheses about cases where humans did not follow the moral law. This is not the proper way to define something, because a definition should include only features that are necessarily connected to the concept. And given that it is impossible to discover the common factor, X'ing, it is impossible to specify the necessary condition for a faculty that can choose for or against the moral law.

In characterizing a Reinholdan *Willkür* as having only the 'liberty of indifference,' Kant criticizes the introduction of this faculty from a somewhat different angle. Reinhold has failed to meet the standards for analysis in terms of faculties, because he introduces a faculty with no account of how it operates.

By contrast, Kant's analysis of the possibility of immoral action avoids both objections. It does not introduce a novel faculty of choice with no basis for choice. Nor does it offer an empirical hypothesis. Rather, it provides a functional analysis that gives the necessary conditions for the possibility of acting in a morally bad way.

Given Kant's clear diagnosis of the mistakes involved in explaining free choice of action through the introduction of a 'free' faculty of choice that executes or contravenes the commands of the moral law, it seems unfair to attribute the view to him. The contrast between *Wille* and *Willkür* has a different purpose. *Wille* in the technical sense of 'pure practical reason,' is not free, because reason is not the faculty of action, desire is—which is why freedom in thought is insufficient to establish freedom in action. For a faculty of desire to be free, however, it must not merely have the properties of a *Willkür*, it must also have the capacity to be moved by the *Wille*. Kant does not explain free action by appealing to a free faculty of choice; he explains free choice of action by appealing to specific kinds of faculties of desire and reason.[26]

6. *How is Morality Possible?*

Kant returns to the issue of the impossibility of explaining how the moral law could be a determining basis of the will in a note in the *Religion* book. He wants to reconcile

the central project of the *Second Critique, viz.,* showing that pure reason is practical, that it can determine the willing, with his claim for the *incompatibility* of determinism and the freedom required for morality in the *First Critique*:

> Die, welche diese unerforschliche Eigenschaft als ganz begreiflich vorspiegeln, machen durch das Wort Determinismus (den Satz der Bestimmung der Willkür durch innere hinreichende Gründe) ein Blendwerk, gleich als ob die Schwierigkeit darin bestände, diesen mit der Freiheit zu vereinigen, woran doch niemand denkt; sondern: wie der Prädeterminism, nach welchem willkürliche Handlungen als Begebenheiten ihre bestimmende Gründe in der vorhergehenden Zeit haben (die mit dem, was sie in sich hält, nicht mehr in unserer Gewalt ist), mit der Freiheit, nach welcher die Handlung sowohl als ihr Gegentheil in dem Augenblicke des Geschehens in der Gewalt des Subjects sein muß, zusammen bestehen könne: das ists, was man einsehen will und nie einsehen wird. (6.50a.)

The resolution is that his incompatibilism does not concern reason determining the will, but the thesis of predeterminism, *viz*., that the state of the world at a preceding time determines what will take place now.

Kant's proposal is a non-starter. Partly through his influence, philosophers characterize 'determinism' in exactly the terms that he uses to describe 'predeterminism.' I offer this text only as evidence that he never gives up on the thesis that morality is possible only because reason (through its moral law) can be, in the terms of the Reflection cited above, not just something that conceives and reasons, but an efficient principle that can occupy the place of a natural cause as a spring of action. The problem is that he also believed that the only efficient causes science recognizes are mechanical and that reasons are not mechanical causes.

Kant was right that reasons operate in a way that is very different from mechanical causation. Although some motions of molecules are no doubt involved, getting a child to understand why the Pythagorean theorem is true or why he should forego acting in a way that he would not want others to copy is very different from transferring a motion. Neither the mathematical understanding nor the restraint from acting is proportional to quantitative features of the impact of the stimuli (i.e. the words) on the sense organs. He is also right that the temporal relations are different. A child may come to understand that he should not do what he would not want others to do and not have occasion to use that knowledge for some time. Further, when he does

use it, he does not use it up: It does not transfer its momentum to the action, thereby dissipating its force.

Where Kant was wrong was in believing that science could only explain phenomena by appealing to laws of motion. By appealing to evolutionary theory we can explain how humans came to have the capacity that he argued was necessary for the possibility of morality: a capacity to have their capacity for action be alterable by reasons independently of self-interest, in particular, by reasoning about whether everyone could do what they intend to do. This approach might also appear to be a non-starter, because it might seem that evolution could not produce a capacity that could lead individuals to act altruistically. As Alan Gibbard argued some years ago, however, it is possible to describe a plausible scenario where humans living in groups could evolve a 'sense of justice' (1982).

More recent discussions have claimed a conflict between evolutionary theory and moral realism. Sharon Street has argued that either a moral capacity is unrelated to moral facts, in which case it would be useless to realists, or it would have to have evolved to respond to moral facts—and such an account is inconsistent with natural selection. Street's attack on moral realism is inconclusive, since, oddly, she considers only natural selection and not the full range of mechanisms (including cultural transmission and developmental constraints) that evolutionary biologists standardly invoke to explain human capacities. More importantly for my purposes, her argument is irrelevant to the project of making Kantian ethics consistent with science.

Kant did not believe that humans had access to alleged moral facts. He explained the possibility of human moral behavior in terms of a set of capacities. In some cases, such as the capacity for restraint in the presence of a stimulus, it was uncontroversial that people had them.[27] Kant also regarded the capacity to alter one's mind through appreciation of principles and the capacity to alter one's behavior through the use of rules of prudence to be obviously present in humans.

If we take an evolutionary or genealogical perspective, we can appreciate one reason why Kant might have found theoretical reasoning and prudential acting so encouraging. The ability to judge independently of alien influences and the ability to act

on prudential 'oughts' involve many of the same sub-capacities as those required by morality, so it is not implausible to think that humans also have the additional capacity required to get them all the way to morality, *viz.* the ability to alter their action-producing faculty of desire through their reason's grasp of moral principles.

Nor is a genealogical perspective foreign to Kant's way of thinking. In the essay on Cosmopolitan History, for example, he tries to explain the possibility of a perfect civil union by offering a genealogy of how it could arise.

There have been at least two other contemporary attempts to make Kantian ethics consistent with science and we can appreciate the special strength of an evolutionary approach by taking a brief look at them. In 1970, Donald Davidson suggested that Kant's problem of reconciling freedom and nature could be solved by adopting the metaphysics of 'anomalous monism' (1970/2003, 207). On this view, although mental types cannot be identified with neurological types—e.g. 'pain' with 'C-fiber firing'—each token mental state is identical with some token physical state (1970/2003, 212). Freedom would be preserved, because there would be no laws connecting mental events *per se*, but determinism would also be preserved, because each token mental event would be identical to a token physical event and that physical event would stand in lawful relation to other physical events (1970/2003, 215). Davidson made only passing references to Kant, but later Ralf Meerbote (1984, 141ff.) and Hud Hudson (1994, 63ff.) tried to work out the connection in greater detail.

To make a long story short, this project probably fails, because it cannot deliver what Kant needs. As Jaegwon Kim, among others, has argued, token identity does not preserve, but threatens to undermine the causality of the mental. Since the physical cause suffices for the physical event, then either the effect would be overdetermined or the mental state would be excluded as a cause, and mental states would be relegated to the status of epiphenomena (e.g. 1983, 54). On this scenario, no reason, including the moral law, could be a cause.

The debate over token identity is ongoing,[28] but evolutionary theory puts a crucial constraint on theories of the relation between the mental and the physical. To see why, consider an obvious example: Vision scientists explain how humans see by appealing to

the structure of the eye; evolutionary theorists explain why human eyes have that structure, by appealing to earlier structures and to the selective advantages conferred by seeing. Neither explanation is more fundamental. A complete theory of vision needs to include both. In the same way, neurophysiology might describe the local or more global features of neural structures that permit humans to alter their behavior by reasoning and evolutionary biology would explain how those structures could be built up from previous structures under the pressure of the selective advantage of rational action. Since *both* theories are required in a complete account of human behavior, an adequate metaphysics must permit reasons as well as impulses to occupy the place of a natural cause. Otherwise, it would be impossible to give the explanation for why humans have the neural structures they do: *viz.,* the structures enabled them to act on the basis of reasons. Thus, the theory of evolution would not only solve Kant's great mystery of how principles or reasons could be causes; it would also avoid the disastrous implication of anomalous monism that the only real or scientifically respectable causes are physical ones.

Starting with Beck (1960, 192), a number of commentators have tried to improve on Kant's solution to the problem of freedom and determinism by turning to the critique of teleological judgment (Meerbote, 1984, 139, Hudson, 1994). The idea is that Kant could have seen that his remarks about the relations between mechanism and teleology provide a different way out of the antinomial conflict. Beck's key move was the suggestion that the principle of determinism not be understood as constitutive, but as regulative: In science, we should always look for causes, but in ethics, we should adopt the agent point of view and act as if the maxim of the will were a sufficient determining ground of the action to be executed or omitted (1960, 193).

Kant's insistence that freedom can only be *proved* from the practical point of view and his striking *Groundwork* assertion that any being who cannot act except under the idea of freedom is really free in the practical respect (4.448) make it very tempting to read him as maintaining that the possibility of morality is secured just so long as humans must think of themselves as free from their perspectives as agents.[29] If this were a sufficient condition for the possibility of morality, however, then there would have been no need to develop the metaphysics of transcendental idealism. Some

scholars who interpret Kant in terms of an agental perspective do so to avoid transcendental idealism. But I think that we can do better. We can reject one of his arguments for transcendental idealism—that it is impossible to understand how the moral law can be an efficient cause—without conceding that morality is chimerical. Kant was, after all, desperate to deny that charge. On an evolutionary account of the development of the capacities required for morality, the capacities would be real and principles would be real causes.

By contrast, as Allison notes, both his dual aspect view and Meerbote's anomalous monism relegate mental causes to the status of fictions (1990, 79). Exactly the same is true for Beck's proposal. Beck's way of saving Kantian ethics from the problems of transcendental idealism has the advantage of drawing on Kantian materials. That cannot be said for an evolutionary approach. The theory of evolution is not a development of Kantian teleology.[30] It is inconsistent with it, because it explains phenomena that he thought had to be explained through purposes, e.g., the eye, without purposes. With the theory of evolution, it is possible to comprehend the various structures of the eye without assuming that an eye ought to be suitable for seeing (20.240). Evolutionary theorists make determinate judgments: the eye has various structures, because those structures and their precursors conferred the selective advantage of better sight. On the other hand, if the goal is to save Kantian ethics, the theory has two enormous advantages: It is true and it provides the theoretical means for showing that Kant's claim that humans are free, because they have an efficacious moral law within, is true—and not merely something that agents must assume is 'true.'

I conclude by considering a likely objection: An evolutionary account undermines the purity of morality, by revealing that the capacity for morality arose through selective advantage. To see how this objection misunderstands the nature of evolutionary explanations, consider an analogous case, that of parental love. Evolutionary theory would explain why parents who loved their children were likely to leave more descendants than indifferent or hostile parents—and hence why parental love is so powerful. It would not show that parents don't really love their children, but are secretly trying to gain some advantage. In a similar way, an account that explained how individuals who could refrain from acting by considering whether everyone could do

what they propose to do increased their representation in subsequent generations would not show that no one really acts out of moral motives. It would show why morality is such a powerful force in human nature.

---

[1] This is not a recent paper of Korsgaard's, but she uses almost the same formulation in her 2009. See p. 75. Further, the idea that what is important to Kant's moral psychology is the capacity of humans to be self-conscious, to stand back and reflect on their desires, has been a central theme of her work. In her widely influential work on the formula of humanity, she argues that it is the capacity to make rational choices that stands behind the value of humanity (see, e.g., her 1986).

[2] That is not to say that there is not a logically possible way to do it: something like Pereboom's Molinist solution work. But that is I think unacceptable, both because it is implausible on its face and b/c it essentially involves a version of pre-established harmony, a position that Kant thought of as intellectual contemptible.

[3] Kant repeats this point in a note in the *Metaphysics of Morals*:

> Denn über das Causal=Verhältniß des Intelligibilen zum Sensibilen giebt es keine Theorie, - und diese specifische Verschiedenheit ist die der Facultäten des Menschen (der oberen und unteren), die ihn charakterisiren (6.438a)

In the Paralogisms Kant says many times that it is impossible to understanding thinking through the resources of materialism. By the B edition, however, he notes that this is no argument for immaterialism of the mental since it is equally impossible to explain thinking with the resources of immaterialism. (B415a)

[4] My attention was drawn to this Reflection by Desmond Hogan's discussion in his handedness paper, note. 56.

[5] The dating of this note is uncertain, with the Academy Edition offering three possibilities, 1778-79, the 1790's and 1776-78. The earlier note is said to be from 1785-88.

[6] In a well-known passage that sets up his attack on compatibilism, Kant tries to argue that the determination of events at one time by events in the preceding time itself establishes that moral principles cannot be efficacious:

> Wenn ich von einem Menschen, der einen Diebstahl verübt, sage, diese That sei nach dem Naturgesetze der Causalität aus den Bestimmungsgründen der vorhergehenden Zeit ein nothwendiger Erfolg, so war es unmöglich, daß sie hat unterbleiben können: *wie kann denn die Beurtheilung nach dem moralischen Gesetze hierin eine Änderung machen und voraussetzen, daß sie doch habe unterlassen werden können, weil das Gesetz sagt, sie hätte unterlassen werden sollen*, d. i. wie kann derjenige in demselben Zeitpunkte in Absicht auf dieselbe Handlung ganz frei heißen, in welchem, und in derselben Absicht, er doch unter einer unvermeidliche Naturnothwendigkeit steht? (5.95, my emphasis)

This argument is invalid. Suppose that someone steals and that all events are determined by preceding events in time. That does not show that moral laws cannot be efficacious, but only that they were not in this case. In other cases, assessments based on the moral law could be the determining factor in the person refraining from thievery. Or if the argument is valid, it is in virtue of an additional premise that judgments based on the moral law cannot be natural causes.

[7] I.e. in the Comments to §6, § 7 and its Corollary.

[8] See my 2011, especially Chapter 9, for a defense of these claims.

[9] See my ms. 2015, for some defense of this argument.

[10] See my 2013 for defense.
[11] Allison (1990, 129) takes what I am regarding as sources of confusions as important additions to Kant's moral psychology.
[12] I'm grateful to Adam Blazej's entry on 'disposition' [*Gesinnung*] for the Cambridge Kant Lexicon for drawing my attention to this fact.
[13] See, e.g., Lawrence Pasternak, 2014, 88.
[14] As Prauss also notes (1983, 93).
[15] I'm grateful to Robert Stern who led me to consider angels in this connection.
[16] In "Empirical Desire" 2014, Allen Wood makes a persuasive case that it is not present desires, but the concern to preserve your options for the future that is the strongest force in self-interest.
[17] Kant denies that this attempt to rationalize away will succeed (5.3), but he must mean that agents cannot deny that they are conscious of moving their faculty of desire through their consideration of the moral law at the very moment when they are doing so. If their faculty of instrumental reason could never put up spurious arguments against the validity of the moral law or their ability to act on it, then they would never prioritize self-love over the moral law.
[18] Although self-deception has often been said to be hopelessly paradoxical, David Pears diffuses much of criticism of accounts that involve self-deception in his 1974.
[19] I have cast the issue in terms of the 'choice' of *Gesinnung,'* but Kant sometimes presents it in terms of a *Hang* or propensity to adopt one or the other *Gesinnung*. Since the moral law is objective and since consciousness of it must move the will of any creature capable of morality, he can explain the difference in behavior between the good and the bad only by appealing to a subjective difference between them, hence the propensity to prioritize the moral law or self-love. As he seems to realize (6.28-29), this is an awkward way to put the matter, because it suggests that people simply have a greater or lesser propensity to prioritize the moral law and that would rule out imputability. The way I have expressed matters does not do justice to the texts involving '*Hang,*' but I am trying to get at Kant's underlying considerations about imputability. A propensity to adopt a *Gesinnung* is imputable only if chosen, so it seems more straightforward just to talk about choosing the *Gesinnung* itself and skip this potentially confusing additional step of choosing a propensity to adopt a *Gesinnung*.
[20] This analysis is also what stands behind his somewhat surprising claim that an individual has had a good or bad *Gesinnung* from 'earliest youth' (6. 25). It is a requirement of his analysis of the necessary conditions for the possibility of moral action that the subject be guided by a maxim that is chosen under those necessary conditions, so as soon as children are thought to be capable of moral action they must also be understood as having a *Gesinnung*.
[21] Gerold Prauss (1982, 70ff.), Paul Franks (2005, 268ff.), and Julian Wuerth (2014, 236ff.) provide illuminating accounts of the course of some of these criticisms.
[22] Beck (1960, 177ff), Allison (1990, 129ff.), Hudson (1994, 151ff.) think that this clarification comes in these late works; Wuerth agrees on the nature and utility of the distinction, but argues that it was implicit in earlier works (2014, 243ff.).
[23] It is far from obvious that Kant thinks of *Willkür* as a faculty or power, despite the standard practice of translating *Willkür* as 'power of choice.' Lin Zhang has noted (in conversation) that *Willkür* is not mentioned as a faculty in the *Anthropology* and as Wuerth (2014, 221ff.) emphasizes, Kant has only three basic faculties, cognition, desire and feeling. Further, as we see below, when he criticizes Reinhold for appealing to an improperly defined faculty he does not use '*Willkür*' but an expression that is explicit about the faculty status, '*Das Vermögen der Wahl* (6. 226). *'Die Willkür*' needs to be distinguished from '*die Wahl*,' meaning a particular choice, but it could simply refer to the whole phenomenon, choice, of which *die Wähle* are instances rather than a power or faculty of executing choices. (I'm grateful to email exchanges with Allen Wood on these translational issues.)
[24] Right before providing his explication of free will Kant seems to claim the opposite relation: *Willkür* is a further specification of a *Wille: "*Insofar as reason can determine the faculty of desire as such, not only choice but also mere wish can be included under the will." 6.213. Here I agree with others (E.g. Allison, 1990, p. ref, Wuerth, 2014, p. ref.) in thinking that Kant uses a broader, generic notion of 'Wille' as well as his particular notion of *Wille* as practical reason.
[25] E.g. Allison, 1990, e.g., 130, Hudson, 1994, e.g., 165.

[26] Although I disagree with Wuerth's view that *Willkür* should be understood as an executive faculty, my view is similar to his in that he lays great weight on the fact that there are three basic faculties for Kant, cognition, desire and feeling (2014, 221ff.).
[27] Evolutionary theories such as Frans deWaal are trying to explain how they could have evolved by looking at restraint in apes.)
[28] The classic reply to Kim was given by Stephen Yablo in (1992). Yablo argues that the relation between the mental and the physical should be understood as that of determinable to determinate and notes that determinates do not exclude their determinables as causes. My appeal to evolution is consistent with Yablo's approach, but has three advantages in the present context. First, it directly answers Kant's concern about how the mental could affect the physical; second it would answer a question that Yablo's account leaves open, viz., how neural states came to be determinates of mental determinables; three, as we see immediately below, it also shows that it is necessary for science and metaphysics to make room for mental causes.
[29] This reading may also be encouraged by Kant's discussion of the prerequisites of thinking and acting in his review of Schultz. There he argues that just as a cognizer can only think under the assumption of freedom—i.e. that his understanding can determine the judgment in accord with objective grounds—an agent can act only under the assumption of freedom—i.e. that the commands of his reason are valid. Although Kant is right that it would be impossible to engage in theoretical or practical reasoning without presupposing some ability to assess facts and values, it does not follow that the agent has the capacity to be moved by values or that the commands of reason are valid.
[30] Hannah Ginsborg argues that there is one sense in which the theory of evolution is consistent with Kant's views. Kant thought that the origin of organisms could not be explained by reference to the basic powers of matter. Evolutionary theory agrees with that position because it traces the 'rise' of any organism to a sequence of contingencies rather than to the basic properties of matter (2015, 314).