

## TONY CHADWICK ESSAY PRIZE 2003 RUNNER-UP

### The Causality of the Will in Kant's *Groundwork of the Metaphysic of Morals*

by Naomi Goulder

To consider how to act an individual must be able to conceive his will as free. In what follows I take this intuitive idea, and look at a specific strand of Kant's argument in the *Groundwork* concerned with the *conditions* of free will. It is important to identify these conditions because in order for the individual coherently to conceive his will as free he is obliged at least to fulfill any conditions upon which he can see that it depends. Not to fulfill them would be to fail to realize a crucial presupposition of his conception of himself as an agent.<sup>1</sup>

What then are the conditions of free will? Kant's project is to show that the primary condition or presupposition of free will is 'a supreme moral law', the idea being that the individual is bound by certain requirements of morality merely in his capacity as a practical being. Such a thesis would diffuse the traditional threat of the amoralist who putatively makes a meaningful choice to act outside of the demands of morality.

There are many strands to Kant's argument. In what follows I concentrate on what I take to be the most plausible of his premises, which opens Chapter 3 of the *Groundwork*:

*'Will is a kind of causality belonging to living beings'*<sup>2</sup>

In §1 I explore this apparently modest and intuitive characterization of the will. In §2 I consider how far it alone could take us towards a moral law that generates the 'strict' duties. Then in §3 I see how far it can take us towards a stronger version of the law, which generates the 'wide' duties.

#### Section 1: Will as a kind of Causality

The intuitive idea driving Kant's above definition of the will is *causation*, in particular some *special kind* of causation. The idea is that we see ourselves as able to choose how to act, and so to interact creatively or freely in the world. In order for the causation of our action to be actively imposed on the world, rather than passively experienced, it must be somehow separate from the *inevitable* causation of the sensible world: 'to be *independent of determination by causes in the sensible world* (and this is what reason must always attribute to itself) is to be free.' For the individual to see himself as more than a mere cog in the great mechanism of nature, he needs to be able to see himself as possessed of his own self-initiated power of causation distinct from the sensible causal network.

In order to explain how such free individual causation could be possible, Kant insists upon a necessary connection between causation and laws<sup>3</sup>: 'The concept of causality carries with it that of *laws* in accordance with which, because of something we call a cause, something else – namely, its effect – must be posited' ... 'Everything in nature works in accordance with laws.' In line with this close relationship, he is then able to identify the distinctiveness of the will's causality in terms of a distinctive feature of those laws on which it is based. Kant says the will is causally free when it acts on the basis of *its own laws* (autonomously) rather than on the alien sensible laws of nature. The will's putative capacity for acting on its *own* laws is supposed to make sense of the intuitive requirement that its causality is *self-determined* rather than determined from outside.

What is involved in the requirement that an individual acts only on the basis of his own laws? Is this any clearer than the obscure notion of 'free will'? Is it of use in the project of finding the ultimate

---

<sup>1</sup> In fact Kant explicitly commits himself to 'the unity of practical and theoretical reason', so in his argument there is also an implication that without free will the individual has no title to reason at all.

<sup>2</sup> All quotations in this essay are from his *Groundwork of the Metaphysic of Morals*.

<sup>3</sup> To the extent that we are motivated to reject the whole Kantian argument we might be tempted to resist this view of causation in terms of extrinsic regularities (laws), and push instead for a simple, intrinsic, causality of the will. I do not however pursue this route.

presuppositions or conditions of such a will? I break the question into two parts: What is it for a law *to be an individual's own* (and so a possible ground of free causal action)? And what is it for an individual *to act on the basis of* such a law? I want to concentrate on the first issue, but before proceeding I briefly explain why the second issue can be passed over as un-answerable.

### ***What is it for an individual to act on the basis of his own law?***

Kant conceives not only causation but also explanation in terms of appeal to laws. To explain something is to cite a relevant law under which it falls. So to ask for an explanation how a law generates an associated action is to presuppose that we can know some antecedently relevant law under which both law and action fall. But *ex hypothesi* in order for the action to be genuinely free the law it is ultimately based on must be the individual's own, and so it is self-defeating to seek some *other* law(s) to explain the necessity:

'reason would overstep all its limits if it took upon itself to explain how pure reason can be practical. This would be identical with the task of explaining how freedom is possible.'

E.g. it would beg the question in favour of determinism to propose a physical law or brain process as *fundamentally* explanatory of how action is generated. It is crucial to the concept of free will that the ultimate law of one's action is one's own, *even though it turns out that how this can be so must be beyond the limits of our comprehension.*

I now return to the first issue raised above:

### ***What is it for a law to be an individual's own?***

Within this idea we can identify *prima facie* constraints both in the requirement of *law-likeness* and in that of *individuality*. If the individual is to see his will as free, he must respect these constraints whatever they entail. I examine each singly in turn, and then look at them taken together.

#### ***Law-likeness***

We know of an individual's subjective principles of action ('maxims') that insofar as they can support causation, they must be law-like. This is simply part of Kant's understanding of causation (above). Laws at least hold universally. So we can derive the requirement that it must be possible for an individual's maxim to hold universally, without exception. The problem with non-universalizable maxims in this context is then not that they are unfair (although they may be), but simply that being non-universalizable they are not candidates for laws and so not possible grounds for causation. If our actions are to be causal, they must at least be based on principles that are universal.

#### ***Law-likeness:***

***act only on a universal principle***

A lot will hinge on how we spell out 'universal'. Further elucidation will be possible once I have addressed *Individuality*, below.

#### ***Individuality***

We are taking as given that in the sensible world the individual is subject to alien laws. We are therefore forced to locate the source of the individual's *own* laws in what Kant calls the 'intelligible' world. The intelligible world is simply *defined* as a world not part of - that is to say, not causally determined by - the sensible world:

'My Idea (of the intelligible world) signifies only a 'something' that remains over when I have excluded from the grounds determining my will everything that belongs to the world of sense: its sole purpose is to restrict the principle that all motives come from the field of sensibility...'

This negative characterization of the individual's freedom from external determination, is accompanied by a positive account about his relation to the intelligible world. The positive thesis is that the individual is able to see himself as participating in the intelligible world in virtue of his rational nature:

'man actually finds in himself a power which distinguishes him... from himself so far as he is affected by objects. This power is reason.'

*Individuality* thus gets spelt out in terms of (two) rationality constraints:

***Rational Accessibility:***

***act only on a principle that is accessible to one's rational nature***

***Rational Applicability:***

***act only on a principle that applies to one in virtue of one's rational nature***

It is natural to suppose that *Individuality* brings with it the further constraint that the particular rational principle must in fact be *chosen* or *endorsed* by the individual who acts on it. But this turns out to be unhelpful because it is not possible for a principle of action to be straightforwardly identified as the description under which the individual *thinks* of himself as acting. As soon as we allow the psychological complexity of subconscious motivations and self-deception, we find that we can't posit a necessary connection between an individual's more or less conscious thoughts and the principles on which he acts. The resulting un-knowability of principles of action has interesting consequences, which I don't go into here.<sup>4</sup> For now I simply draw the conclusion that we must leave the three constraints of ***Law-likeness***, ***Rational Accessibility***, and ***Rational Applicability*** in their minimal formulations, without attempting further psychological elucidation or content.

We have seen that free will depends upon a special kind of causation, and that this in turn depends upon the individual acting on his own laws. We have identified three constraints inherent in this idea of autonomy, and can make the conditional claim that if an individual does act freely then it is on the basis of a principle that respects these constraints. Now we can ask, what is involved in these constraints which are apparently presupposed by our practical natures? Again, are they any clearer than the obscure notions of free will, of individual causation, or of acting on one's own laws? In §2 I try to spell them out along Kant's lines in the form of tests that the individual can apply to potential maxims. I then look at how far these tests take us towards something like a moral law.

## **Section 2: Conceptual Universalizability & Narrow Duties**

I examine each of the three constraints derived above:

***Rational Accessibility***

Without strictly intelligible content, a maxim cannot be grasped by the individual's purely rational nature. The *rational accessibility* constraint therefore generates the requirement that a maxim have coherent propositional content:

***Propositional Coherence test:***

***act only on a strictly coherent proposition***

---

<sup>4</sup> If an individual can act on the basis of rational laws without actually articulating those laws to himself, then his actions can be causally free independently of his conscious intentions. Would we really want to say in those cases that his will was free (really his own)? If we can't relate rational principles of action more closely to an individual's thoughts then intelligible rational laws begin to look just as alien as the sensible natural laws we have been trying to avoid. Moreover, how are we to distinguish between an individual's action being *based on* reverence for a rational law and its merely *falling under* that law? This is important because, if we accept Kant's inference from free will to a supreme principle of morality, there could be an argument from our apparently evident free will to the existence of some strictly moral principle of action in the background of every apparently immoral act. In this case an individual could seem to act immorally, but it would always turn out either that some moral motive was the true spring of his action, or else that he didn't act freely in the first place. Either way it becomes hard to make sense of moral responsibility.

This test on its own does not take us far because a rationally accessible principle could have *any* propositional content. It permits, for example, the paradigmatically immoral principle ‘Lie when you want to’. Indeed, it might be thought that any maxim that the individual is able to consider at all must already as such respect a minimal *propositional coherence*. This is where the importance of the other constraints comes in. It is precisely when we add a universalizability test generated by *law-likeness* and *rational applicability* that an incoherence emerges in the maxim in question: *There is no fully worked out proposition to the effect that ‘every’ rational being lies when he wants to; on analysis the ‘universalized’ sentence turns out to be meaningless.*

The *propositional coherence* test seems to me uncontroversial. The crucial *universalizability* test is what needs to be argued for. As it happens conceptual universalizability is in a sense over-determined, because it is entailed both by *law-likeness* and by *rational applicability*. Kant does not make this over-determination explicit, and in the end the two constraints are intimately connected. Nevertheless for completeness I spell out the universalizability test in terms of both of its apparently divergent motivations.

### ***Law-likeness***

The universality of any law is relative to the terms of that law. If the law says ‘*as are bs*’, then universality attaches to the *as* and the *bs* so that in order to be a genuine law *all as* must be *bs*. In the current situation, the rational applicability constraint restricts one of the terms of the law to *rational beings*. *Law-likeness* then entails that whatever maxim is supposed to apply to one rational being must thereby be supposed to apply to all:

*Universalizability test:*

*act only on a maxim that could apply to all rational beings*

### ***Rational Applicability***

This final constraint states that the individual should act only on a principle that applies to him in virtue of his rational nature alone. Without this *rational applicability* restriction, a merely universal principle could contain a relativizing principle within it, so as to be formally universally applicable but to *apply* only vacuously to some rational beings – e.g. the maxim ‘Lie if you want to, if you are over 6ft tall’. In this example, the maxim’s *application* is dependent on contingent sensible features of the agent, and so fails to meet the *rational applicability* constraint which states that a maxim must apply to rational beings *as such*. Here we have another motivation for a universalizability test: One way of knowing if a maxim is applicable purely in virtue of an individual’s rational nature is to consider if it would apply to him *whatever his sensible nature*, which can equivalently be formulated as a test whether it would apply to every possible rational being (i.e. rational beings with infinitely various sensible natures):

*Universalizability test:*

*act only on a maxim that could apply to all rational beings*

The disparate motivations of *law-likeness* and of *rational applicability* both generate an identical universalizability test. We can now see that a distinctive moral content or condition on free will is straightforwardly derived through the *combination* of propositional coherence (via *rational accessibility*), with universalizability (via *rational applicability* and *law-likeness*). This *overall* condition can be formalized as a single test:

### **(CU) CONCEPTUAL UNIVERSALIZABILITY TEST:**

*act only on a maxim that could coherently apply to all rational beings*

This is a formula with definite moral content. (CU) is indeed equivalent to Kant’s weak formula of universal law:

‘act only on that maxim which could at the same time *be a universal rational law*’

According to the argument so far, respecting (CU) really is a minimal requirement for being entitled to conceive ourselves as free agents. This is because our freedom depends upon our actions springing from our own (non-alien) laws, and only such maxims as accord with (CU) are possible candidates for such laws. (CU) generates precisely what Kant calls the class of ‘strict’ duties. The strict duties are demarcated as the negations of those ‘actions (that) are so constituted that their maxim cannot even be conceived as a universal law of nature without contradiction’.<sup>5</sup> This on its own is a substantive result. It remains to be seen, however, if the argument from free will can take us any further, towards justifying the ‘wide’ duties.

### Section 3: Volitional Universalizability & Wide Duties

Kant’s ‘supreme principle’ (VU) determines not only the strict duties but also what he calls the ‘wide’ duties. Wide duties are the negations of those in which ‘it is... impossible to *will* that their maxim should be raised to the universality of a law of nature, because such a will would contradict itself’. An example of a wide duty would be to help one’s neighbour when he is in need. It is important to the moral project to see if our ‘free will’ premise can generate (VU), because in practice the derivation of most of our actual moral precepts depends upon this volitional version of the universalizability test:

**(VU) VOLITIONAL UNIVERSALIZABILITY TEST<sup>6</sup>:**  
***‘act only on that maxim through which you can at the same time will that it should become a universal law’***

Can the volitional universalizability constraint be derived from our initial intuitive premise, with its driving intuitive idea of individual causation? (CU) *seems* to be part of the thought from which Kant generates this ‘supreme principle of morality’, and yet on its own (CU) entails no more than the *strict* duties. This raises the worry that in order to rationally justify the wide duties we will need an additional premise – possibly less obvious, and more susceptible to sceptical attack. (Of course, it may be that once we have antecedently identified the wide duties, we find that a convenient test for them looks something like (VU) – but that is a different matter.)

To go back briefly to the case of conceptual universalizability and strict duties: The fact that we don’t in fact all regularly lie means that the institution of lying can persist, and the individual can individually lie. Nevertheless he cannot do so freely, because causation depends upon universal laws, and free individual causation depends specifically upon acting on universal laws that apply to rational beings as such - whereas the maxim of lying cannot be coherently universalized as a law across all rational beings.

Kant’s justification for (VU) and the consequent derivation of wide duties depends not on anything inherent in *conceptual universalizability*, but on a special emphasis upon something like what the

---

<sup>5</sup> Before going on to look at where (CU) falls short of (VU), I look briefly at the general connection Kant draws between laws and duties. In so far as an individual is rational *as such* his subjective principles of action will naturally accord at least with (CU). It is part of conceiving oneself as practical that one’s maxims accord with (CU), so it is not an imperative from without but a presupposition of one’s own practical rational nature. However, human beings are not *merely* rational, they also have a sensible, externally determined, nature. If the individual does not use his practical reason, or alternatively if he reasons coherently up until the last moment and then out of tiredness or evasion or something else he lets his sensible impulses take over, then his subjective principles of action may not be chosen *on the basis of* (CU). In such cases Kant says that although the individual *may not in fact* act on the basis of (CU), nevertheless in his capacity as a practical being he *ought to have* done so. That is, so far as he had reasons for acting at all he had an over-riding (categorical) reason for respecting (CU). This is clearly an internal reason. (CU) generates the class of strict duties because the strict duties are simply defined as the negations of those that do not respect (CU).

Kant’s argument implies that all failures to respect (CU) are cases of weakness of the will, which might seem to bring with it the idea of a kind of malfunction or mistake. In another essay I’d like to look at whether Kant can adequately accommodate respect for persons in their capacities as sensible beings (beings who identify strongly with many of their admittedly determined physical and emotional characteristics, and perhaps identify with them specifically *because of* their sensible nature –love for one’s close blood relations, for instance).

<sup>6</sup> My title (Kant’s is ‘the formula of universal law’, but this title doesn’t bring out the distinctive features I want to discuss.)

individual would overall want to universalize. He writes of a contradiction in the individual's will as though there is an issue of literal impossibility, and *ex hypothesi* the only constraints on a rational being as such are purely rational constraints. But there is no controversy over the fact that the negations of the wide duties do indeed respect the internal rational constraints of (CU). The rational contradiction that Kant insists on is thus not an incoherence inherent in the maxim as such, but rather a contradiction between the content of the admittedly coherent maxim, and the content of some *other* rational principle also supposedly subscribed to by the individual.

If an individual's maxim expresses a universal law of the form 'all *as* are *bs*', another principle held by that individual could generate a contradiction by expressing an individual instance of the form 'this *a* is **not-*b***':

'the contradiction (is) that a certain principle should be objectively necessary as a universal law and yet subjectively should not hold universally but should admit of exceptions.'

To put this in the context of the example I've been considering, the contradiction would be between the coherent maxim:

'if *x* is rational, *x* doesn't help his neighbour'

and the individual's coherent particular desire that his own neighbour John sometime (rationally, freely) help *him*:

'John is rational, and John on some occasion helps his neighbour (me)'

There is indeed a literal contradiction here between the law of the maxim (not helping one's neighbour) and the law embedded in the particular desire (helping one's neighbour). And this contradiction is not removed if the particular desire for help from one's neighbour relates only to some future time or alternative sensible state of affairs, because the law embodied in the maxim is *ex hypothesi* (by (CU)) absolutely universal across time and space, and totally independent for its application of such sensible factors.

Of course, it is not contradictory for rational beings *as such* to will maxims like 'don't help your neighbour'. We can consistently imagine rational beings who prefer not to be helped, or perhaps a physical environment in which humans themselves prefer to live self-sufficiently, so it is clear that a law to the effect that no-one helps his neighbour is impossible only conditionally on our sensible situation. Nevertheless, *given* our sensible situation and the particular desires it elicits, it *seems* to have been shown that we cannot freely act on the negations of the wide duties any more than we can on the negations of the strict ones. I want to question this.

We saw in §2 that in order to act freely, the individual must act on the basis of some coherent universal rational law. This generates the class of strict duties. We now *seem* to have shown that in order to act freely, the maxim on which this individual acts must not rationally conflict with *other* laws or particular principles to which he subscribes. But is this quite right? It is true only to the extent that for free action the individual needs to thoroughly integrate all his various propositional attitudes. Kant writes of such failures of integration of principles and desires, etc. as entailing *contradictions in the individual's will*, but these failures are only relevant or problematic if the individual cannot act on the basis of anything less than his whole person. Perhaps if the individual is able to respect a certain amount of dis-integration within himself, the special causality of the will can work piecemeal.

It appears that to get from the *conceptual universalizability* test (CU), to the stronger *volitional* version (VU), Kant requires an additional constraint to the effect that a law or maxim is only an individual's own if, as well as being *individually rationally accessible* (internally propositionally coherent), it also forms part of a complete set of principles which are *mutually rationally accessible* (all consistent with each other):

***Integrated Coherence:***

***act only on a principle that is consistent with all one's other principles***

This *integrated coherence* constraint is extremely plausible - so much so that Kant does not even argue for it. But while *integrated coherence* is of course inevitable for rational beings *as such*, I think it may be too severe and psychologically unrealistic to be relevant *even as an aspiration* for human beings.

The individual human is involved in diverse projects and commitments and ways of life, which inevitably give rise to an irreducible multiplicity of incommensurable beliefs, desires, and values. No principle could be of interest which required him to give up the diversity of the very fabric of his life. If free action really requires strict rational integration of propositional attitudes, then at best free action will have to be relative to mere 'aspects' of an individual's personality; the *will* will have to operate on the basis of laws which form only a mere subset of the individual's total character.

A certain amount of dis-integration of the individual's personality is inevitable regardless of moral issues. For this reason I do not see the will as *distinctively* frustrated by the admitted specific contradiction between on the one hand the maxim of not helping one's neighbour, and on the other hand the particular desire to be (freely) helped by him. *The individual can identify himself with the maxim at one moment, and identify himself with the particular desire at another moment – thereby achieving free action in each instance.* There is thus a literal sense in which one who freely flouts the wide duties thereby acts *without integrity*. This sounds like a trick, and there is definitely a problem with it; but I think dis-integration is *already* an unavoidable commonplace of practical deliberation. The contradiction in willing that Kant identifies is not specific to the negations of wide duties. His claim that such contradictions frustrate freedom therefore *pervasively* calls into question the possibility of free human action.

In a future project it would be interesting to look in more detail at the ways in which incommensurable propositional attitudes can render rational integrity impossible, and how far this lack of integrity can threaten individual freedom.