

The moral law as causal law

Robert N. Johnson

A well-known gap exists in Kant's argument that the categorical imperative is the fundamental principle of morality. The gap is between the claim that rational agents act only on maxims that could serve as universal laws and the more substantial requirement contained in the categorical imperative.¹ There does not appear to be any obvious route from the seemingly trivial requirement to conform to universally valid laws to the controversial and substantive requirement of acting only on maxims that one can at the same time will that all other rational agents act on.² Kant apparently assumed that we conform our wills to universally valid laws only if we act on maxims that we can will to be universal laws, but many have pointed out he gives no defence of this assumption. In what follows, I focus on a preliminary step to connecting the two claims, the step that connects rational agency to conformity to universally valid laws. Why must the will of a rational agent, insofar as he is rational, be such that it is universally valid in any sense? My hope is that by first considering this question, some light can be shed on the nature of the remaining argument for the categorical imperative.

Many readers assume that for Kant, a rational agent must conform to universally valid laws because rational agency is *rational*, and rationality requires this sort of universal validity. For instance, this view appears to be implicit in Onora O'Neill's claim that 'the interest of a Kantian universality

I want to thank Stephen Darwall, Thomas E. Hill, Jr, Matthew McGrath, Andrew Melnyk, Mark Schroeder, Jens Timmermann, Alexander von Schönborn and an anonymous referee for helpful comments and discussions of the ideas in this chapter.

¹ Henry E. Allison, 'On a Presumed Gap in the Derivation of the Categorical Imperative', *Philosophical Topics* 19(1) (1991), 1–15; Bruce Aune, *Kant's Theory of Morals* (Princeton University Press, 1979), pp. 28–31; Thomas E. Hill, 'Kant's Argument for the Rationality of Moral Conduct', *Pacific Philosophical Quarterly* 66 (1985), 3–23, also in P. Guyer (ed.), *Kant's Groundwork of the Metaphysics of Morals: Critical Essays* (Rowman & Littlefield, 1998); Samuel Kerstein, *Kant's Search for the Supreme Principle of Morality* (Cambridge University Press, 2002); Christine Korsgaard, 'Kant's Formula of Humanity' in her *Creating the Kingdom of Ends* (Cambridge University Press, 1998), pp. 106–32; Allen Wood, *Kant's Ethical Thought* (Cambridge University Press, 1999), p. 81.

² See Hill, 'Kant's Argument for the Rationality of Moral Conduct', 268.

test is that it aims to ground an ethical theory on notions of consistency and rationality rather than upon considerations of desire and preference.³ This understanding of Kant's view as grounding universalization in rationality, in turn, draws objections from, for instance, particularists who believe that rationality is compatible with non-universalization.⁴ But, as I contend below, the requirement that rational agents must conform their wills to universally valid laws is not based on the fact that rational agency is *rational*, but on the fact that it is *agency*. Indeed, the argument leading to the first formulation of the categorical imperative relies on the idea of rational willing as a kind of causation in order to show this. The moral law is a unique causal law, and it is in virtue of this fact that the moral law requires one's will to be universally valid.

My plan is as follows. I start by discussing the rationale behind thinking that conformity to universally valid laws is a requirement of rationality, where I conclude that it is at least an open question what grounds the 'universal' in the formula of universal law of nature. I then explain the role that conformity to universally valid causal laws plays in Kant's project of seeking out and justifying the supreme principle of morality. I follow this with a discussion of some burdens this leaves for Kantians who favour the first formulation of the categorical imperative as a criterion or test of right action. I end by claiming – admittedly without much argument – that while reason does not provide the universalization requirement in the categorical imperative, it does provide a spontaneity requirement. Whatever the moral law is, it must be the law of spontaneous self-governed agency, because our agency is the agency of rational beings, and reason is a spontaneous faculty. This is, of course, just the familiar message of Kant's ethics, that the moral law is the law of an autonomous will. Thus, the innovation I introduce to understanding the argument for the first formulation in what follows is that the universality required by the moral law originates in the causality of the will, while the spontaneity of the will comes from the rational origin of that causality.

A preliminary methodological point is worth making. The argument of the *Groundwork* is, by anyone's standards, exceedingly difficult to extract from the text. That is because, as is plain, the argument is very compressed and the writing is not particularly elegant. So it is often not obvious how to

³ Onora O'Neill, 'Consistency in Action' in Paul Guyer, *Kant's Groundwork of the Metaphysics of Morals: Critical Essays*, pp. 103–31, at p. 127; see also Hill, 'Kant's Argument for the Rationality of Moral Conduct'.

⁴ See, e.g. Jonathan Dancy, *Moral Reasons* (Oxford: Blackwell, 1993), p. 83.

move from one claim Kant makes to another. As a result, there are a variety of different, but reasonable, ways of putting together these claims into an argument. In what follows, I piece together a variety of texts in a way that makes philosophical sense overall, but I make no claim that this is the only defensible way of doing so. My main aim is to offer an argumentative structure that I believe makes sense out of what Kant says, and although I do claim that my reading raises problems for the defence of the first formulation of the categorical imperative, it does so by solving other problems.

I RATIONALITY AND UNIVERSALITY

A rational agent, insofar as she is a rational agent, acts on maxims that conform to principles to which all other similarly situated rational agents conform. Unfortunately, it is perfectly consistent with acknowledging that other similarly situated agents would be rational to adopt her principle that, nevertheless, a rational agent could not will that other agents adopt it. She need not be able to will this, and may well will that others do not adopt her principle, though she acknowledges that they could adopt it. That this does not follow leaves a gap in the argument establishing the first formulation of the categorical imperative. Possessing only the fact that rational agents conform their wills to what would apply universally in this sense, we cannot conclude that they must be able to will that everyone adopt their maxim. But why must a rational agent's volitional principle be universal in any sense?

It seems that many of Kant's readers think that it is in virtue of their rationality that it must be universally acceptable, willable, or whatever. Henry Allison, who regards this as only the 'first step' in a Kantian argument for universalizability, nicely summarizes this way of reconstructing the connection between rationality and universality:

In claiming that one's reason for acting in a certain way is 'good' in the sense of justifying reason, one is, implicitly at least, assuming its appropriateness for all rational beings ... Since ... to regard one's reason for acting in a certain way as good is to assume its legitimacy for all rational beings in similar circumstances, it would seem, so the argument goes, that rational agents cannot reject the universalizability test without, at the same time, denying their rationality.⁵

The idea is that it would be irrational to hold that one's reasons justify one's action and yet deny that it would be appropriate for all other rational beings

⁵ Allison, *Kant's Theory of Freedom*, pp. 204–5.

in similar circumstances to act on those reasons as well. So rational agents must hold that the principle of their wills is appropriate for any rational agent to adopt. One passage from the *Groundwork* that might be thought to support this reconstruction tying universal validity to rationality is this (*G* IV 420–1):⁶

When I think of a *categorical* imperative I know at once what it contains. For, since the imperative contains, beyond the law, only the necessity that the maxim be in conformity with this law, while the law contains no condition to which it would be limited, nothing is left with which the maxim of action is to conform but the universality of law as such.

Rational agents, at least insofar as they are rational agents, conform to categorical as well as hypothetical imperatives, or so is the message of the *Groundwork*. And while you can have no idea what taking the means to your ends will require of you until you have some end in mind, you do not require an end in order to know what a categorical imperative requires of you, and so you know ‘immediately’. Categorical imperatives tell us to conform our maxims to practical laws without any such ‘condition’. But, the proposed reconstruction of this argument goes, practical laws (as Kant explains in the footnote associated with the term ‘maxim’) are objectively valid. And objectively valid laws are ‘valid for every rational being’. Therefore, rational agents act on universally valid laws (*G* IV 421 fn.). We might say that the proposed connection between rational agency and universality is through objectivity: rationality brings with it objectivity (in justification) and objectivity brings with it universal validity.

However, Kant himself does not say that rational agents must act on universally valid principles *because they are rational*. It may well be, of course, that their rationality demands this. But it may also be that rational agents are bound to universalize because being an *agent* of a certain sort demands it. There is no indication in the above passage from Kant which of these is so. However, as I will argue below, I believe the answer has in fact already been given in the discussion leading up to it. That discussion is best understood as claiming that it is because rational agents are causes that they must operate in a universally valid way. I return to explain why I think this prior discussion should be understood in this way below. Before I do, I want to say a few more things about the connection between rationality and universality.

⁶ See, e.g. Wood’s gloss on this passage in Wood, *Kant’s Ethical Thought*, pp. 78–82.

Another route connecting rational agency to universal laws is through consistency rather than objectivity:

- (1) Rationality requires consistency.
- (2) Consistency requires conforming to universally valid laws.

From (1) and (2) it does indeed follow that rational agency requires conforming to universally valid laws in virtue of its being *rational*. It would be inconsistent, and hence irrational, to regard a principle as justifying your doing some action, but deny that it would justify similarly situated agents. Of course, rationality in this sense also requires conforming to the laws of non-contradiction, excluded middle and other laws of logic. It would, for instance, be contradictory, and hence also irrational, to adopt a self-contradictory principle. Rational agency thus brings these laws of reason that are valid for every rational being to bear on actions.

The difficulty with extracting universal validity out of rationality in this way is that it puts the moral law on a par with logical laws. But if the moral law is on a par with logical laws, it is difficult to see how it is supposed to be a law of *practical* rather than *theoretical* reason. The analysis of the concept of a rational being yields the conclusion that rational beings conform to logic. So if the conformity with universal laws contained in the categorical imperative is on a par with the laws of logic, it would seem the categorical imperative is a law of *thought*. But while rationality requires thought be one way rather than another, the puzzle is how to use this fact to explain how rationality requires action to be one way rather than another. More precisely, a straightforward application of logical laws to action would seem only to yield principles it would be impossible not to fulfil such as 'For every action, either do it or don't do it' or 'For every action, don't both do it and fail to do it'. Since the logical laws are the laws of what is logically possible, they have nothing to say about what one ought to do.

R. M. Hare, who took himself to be at least in part a follower of Kant, got conformity to universal laws out of rationality through the notion of *supervenience*.⁷ Rational agents (oversimplifying) treat like cases alike. And to treat like cases alike is to do such things as the following: if I judge I ought to \emptyset in C_1 , then for every C_n that is like C_1 , I should judge I ought to \emptyset in C_n . If I am in circumstances that are exactly like other circumstances I was in and judged I ought to perform some action, then I should not judge anything other than that I ought to perform that action in the present circumstances. But to judge that in all C_n that are like C_1 I ought to \emptyset is just to judge that I

⁷ R. M. Hare, *The Language of Morals* (New York: Oxford University Press, 1991), p. 70.

ought to conform to a universally valid law. In essence, reason requires us to be *non-arbitrary* in our judgements about what we ought to do.

Hare's supervenience requirement is, however, in one sense weaker and in another sense stronger than Kant's. It is weaker in the sense that all that it requires is that, for all rational agents, if one judges *x* is *F*, then one must judge anything else that is like *x* is also *F*, and that, as Hare himself admitted, does not rule out a Nazi morality.⁸ Indeed, human beings, even idealized ones, will differ in their ends, and given this, supervenience will not yield universally valid prescriptions.⁹ Rationality may require supervenience, but supervenience does not provide laws valid for every rational agent. Hare's supervenience requirement is in another sense, however, stronger than Kant's, since for Hare universalization requires us to take each person's interests into account equally. But, quite famously, Kant's universalization does not do so, since the impact of an action on people's interests does not play a foundational role as it does for Hare.

Perhaps Kant thought that rational agents conform their wills to universally valid laws because he thought that rationality brings universality. But there is room to doubt that this is the source of the universalization requirement. Consider a very general point about the idea that a law or rule is 'valid of' or 'applies to' someone. Take some set of persons *S* and some law *L* that governs the actions of a person *x* who is in *S*. Knowing that *L* governs those who are in *S* requires knowing both *L* and the property *F* in virtue of which that *L* applies to those who are in *S*. That is, *L* applies to *x* who is in *S* because all *S*s are *F*. Thus, answering the question 'Does *L* apply to *x*?' requires us to know what the *F* is in virtue of which *L* applies, and the content of the set of laws that govern persons with that property.

For instance, suppose I want to know whether the rule 'Be thrifty' applies to a given group of people. I will have an answer once I discover that the members of the group are Boy Scouts and that all Boy Scouts are adherents of the Scouting Laws. It is in virtue of being an adherent to the Scouting Laws that the law 'Be thrifty' governs Bobby, a given Boy Scout. Naturally, not every person who is actually a Boy Scout is also an adherent of the Scouting Laws, but let's suppose 'Boy Scout' is like 'rational agent' in the sense that someone who is a Boy Scout conforms to that rule, insofar as he is acting as a Boy Scout, or perhaps a Boy Scout who violated that rule is not in some sense being a *real* Boy Scout.

⁸ R. M. Hare, *Freedom and Reason* (Oxford University Press, 1963), p. 192.

⁹ See David Sobel, 'Do the Desires of Rational Agents Converge?', *Analysis* 59 (1999), pp. 137–47, for a treatment of this question, in connection with the views of Michael Smith.

Suppose now that Bobby is also a member of T, another set of persons, and because of this is not only an F but, in virtue of his membership in T, also a G. It then becomes a live question in virtue of which property L applies to Bobby, F or G. Suppose Bobby is both in the Boy Scouts and is a member of a martial arts group, say, Tae Kwon Do. Members of Tae Kwon Do are adherents of the Tae Kwon Do goals. We can ask in virtue of which membership the rule to be thrifty applies to Bobby. Is it in virtue of being a member of the Boy Scouts or Tae Kwon Do that Bobby must be thrifty? And, of course, there are three possibilities: he must be thrifty either because he is a Boy Scout, an adherent of Tae Kwon Do, or both.

Likewise, if there are laws or rules that apply to every rational agent, they will apply in virtue of the fact that every rational agent is rational, or is an agent, or both. More precisely, the rules that apply to rational agents will be such that their nature and content will be determined in part because of their rationality and in part because of their being agents. We could, for instance, imagine rational beings that are not agents, beings that do not bring about events in the world.¹⁰ Whatever rules or laws apply to them would be distinctly different from those applying to rational beings who are also agents, and, moreover, whose agency operates through their rationality. So it is an open question whether the categorical imperative, as the supreme law of rational agency, requires universalizable maxims in virtue of the rationality of rational agents, their agency, or both.

It doesn't follow from this, of course, that Kant didn't, in the end, think as Allison or Hare does about these matters. The foregoing only shows that it is not obvious that this is so, or at least whether Kant thought in these ways about universalization. Thus, it is worth exploring an alternative proposal. In summary, my alternative is this: if reason is a cause, then at minimum it must act as a cause. And to act as a cause is to conform to universal law, or so acting as a cause is on Kant's view.

2 ROLE OF THE IDEA OF CAUSATION IN IDENTIFYING AND JUSTIFYING THE CATEGORICAL IMPERATIVE

The first paragraphs of Section III of the *Groundwork* show Kant arguing that conformity to universal law by rational agents is because they are causes, not because they are rational. There he asserts these two claims (*G* IV 446):

¹⁰ Kant himself imagines such beings at *Religion* VI 26 fn.

- (a) 'Will is a kind of causality of living beings insofar as they are rational'
 (b) 'The concept of causality brings with it that of laws in accordance with which, by something we call a cause, something else, namely an effect, must be posited.'

(a) and (b) imply that it is in virtue of being a cause that a free rational agent must conform to universally valid laws. Thus, causation, not rationality, brings with it the idea of universally valid laws. Rational agents are causes and so are thereby bound by universal laws. And, indeed, Kant draws just this conclusion (*G IV* 446):

- (c) 'So freedom, although it is not a property of the will in accordance with natural laws, is not for that reason lawless but must instead be a causality in accordance with immutable laws but of a special kind.'

They must be of a 'special kind' because no *natural* causal law can direct the workings of a free will; the causality of a rational agent must be non-natural, a causality of which rational agents, simply in virtue of their being rational agents, are the sources. Thus, the claim that an agent is a cause, together with the contention that causes do their work in conformity with universally valid laws, delivers the conclusion that rational agents, because they are agents, do their work in conformity with universal laws. The additional step that nature cannot be the source of the law according to which the causation of a rational agent is exercised allows us to conclude that this law is one and the same as the law of an autonomous will. Kant goes on to explain that whether rational beings such as ourselves *are* free causes in this sense is a 'third cognition', making the claim that we as free rational agents are bound by universal laws a synthetic claim (*G IV* 447).

Notice an important consequence of thinking of universalization as the upshot of causation rather than rationality. Every rational agent, Kant states, must act under the Idea of freedom. To act under the Idea of freedom is to act under the idea that one is a cause of a special kind, a 'first cause' of one's actions, but nonetheless a cause. If causation brings with it universal law, then in order to act under the idea that one is a (first) cause, one must act under the idea that one's actions are the upshot of one's will *in an essentially repeatable way*. It cannot be that one bears, in virtue of one's causality, a *sui generis*, one-off relation to one's action, or at least one cannot act under the idea that one bears such a relation. That is, acting under the Idea of freedom commits rational agents to acting under the idea that the law under which their volition falls is a causal law covering every other rational agent's causality. Rational agents are thus committed to this, not because they are rational, but because the alternative is that they would not be a self-originating cause of their actions.

Kant makes repeated references to the moral law as ‘a law of causality through freedom’ throughout his ethical works, and elsewhere in the *Groundwork*.¹¹ I want to discuss one of these references to which I referred above, the argument in *Groundwork* II leading up to the statement of the first formulation of the categorical imperative. That argument concerns how imperatives are possible. The line of reasoning hinges on the idea that willing is a kind of causation. This is why Kant held it to be analytic that whoever wills an end also wills (insofar as he is a rational agent) the necessary means to that end (*G* IV 417):

In the volition of an object as my effect, my causality as acting cause, that is, the use of means, is already thought, and the imperative extracts the concept of actions necessary to this end merely from the concept of a volition of this end.

This idea introduces a feature of causation on Kant’s view. Very roughly, if *x* is the cause of *y*, then if there is some *z* that is a cause of *y*, then either *x* is the cause of *z* or *x* = *z*. Thus, if one is to be the cause of an end, one must be the cause of whatever are the necessary means to that end. So in thinking of oneself as a cause of an end, one is *thereby* thinking of oneself as a cause of the means by which that end is to be brought about. We derive this law according to which our causality as rational agents operates – the law of willing the means to the ends that we will – from simply the idea of being the cause of an end.

By contrast, proving that rational agency demands conformity to the categorical imperative is not as easy. We cannot derive the law according to which our causality as rational agents operates in willing *moral* actions analytically from the idea that we aim to be the cause of some end to which the action is a means. Instead, the idea of ourselves as a cause is connected ‘immediately with the concept of the will of a rational being as something that is not contained in it’ (*G* IV 420 fn.). The idea of the law according to which a rational agent is the cause of actions that are *moral* actions is not ‘contained in’ the very idea of an agent who (contingently) wills some end, since it is not supposed to be the fact that we will be a cause of such an end that is the ground or reason for our willing the action. We will the action for its own sake, not for the sake of the end it brings about or realizes. Nevertheless, this law is ‘immediately’ (*a priori*) connected with rational agency. This entire line of reasoning moving toward a statement of the fundamental principle of morality thus relies on the idea that, whatever

¹¹ See, e.g. *CpVV* 47; also *V* 42–58, *V* 103–5, *passim*.

role rationality ultimately plays, it will also be the nature of causality that constrains what the fundamental principle of morality turns out to be.

Again, a key move in the line of reasoning in *Groundwork* I identifying the categorical imperative as the fundamental principle of morality relies on viewing the issue as having to do with the nature of causality in rational agency. That move is the argument for the premise that duty is conformity to universal law as such. This argument appeals to the nature of the motive of a good will, respect for duty. Respect, Kant argues, cannot be taken up toward something that gets a grip on us by way of the effects of our actions. This is because the object of respect must satisfy a kind of uniqueness condition: for any effect of our action, that effect (*G IV 401*):

could have been also brought about by other causes, so that there would have been no need, for this, of the will of a rational being, in which, however, the highest and unconditional good alone can be found.

Respect is thus elicited only by something *unique*, in particular, a unique *cause*, while anything that is caused by a will of that sort could have been caused by something other than it. If true, this claim has the consequence that respect can focus only on the ‘activity of the will’ itself, insofar as it is an activity of a cause, and not on *what* the will causes or the outcomes of agency (*G IV 400*). Respect for duty thus involves conceiving of the special way in which practical reason operates as a cause. In particular, it represents the causation of practical reason as unique. And in turn, that, Kant states, is ‘the representation of the law in itself, *which can of course occur only in a rational being*, insofar as it and not the hoped for effect is the determining ground of the will’ (*G IV 401*). The ‘it’ in the last clause is the representation of the law. Why it can ‘of course’ occur only in a rational being is explained later in the *Groundwork* II: only a being with reason can derive actions from laws. But if I am right, it is in representing practical reason as a cause that we thereby are representing a law, but a unique one that can only be found in a rational will.

In *Groundwork* II, Kant spells out more fully the line of argument from agency to universalization, although the reasoning here, well known as it is, is also quite compressed. The operation of a rational will is itself, he argues, the source of the requirement of conformity to universal laws. He sets up the discussion by stating that, in seeking out the content of the moral law (*G IV 412*):

we must follow and present distinctly the practical faculty of reason, from its general rules of determination to the point where the concept of duty arises from it.

This line of reasoning would make most sense if Kant believes that a correct understanding of reason insofar as it is a cause will itself reveal the universalization requirement. The discussion that follows turns to the will's 'general rules of determination', that is, the rules according to which it brings about action (read: causal laws). The determination of the will is, he states, according to *representations of laws* and not merely, as with everything else in nature, according to laws of nature. His first claim is thus (G IV 412):

Everything in nature works in accordance with laws. Only a rational being has the capacity to act *in accordance with the representation* of laws, that is, in accordance with principles, or has a *will*.

If rational beings alone had the capacity to entertain representations, then of course *a fortiori* only rational beings would have the capacity to act in accordance with their representations of laws. But having the capacity for representation is too broad a feature to explain what it is to be a rational agent, and, anyway, animals also have the capacity to entertain representations. One might suppose, instead, that Kant's idea is that other beings, while capable of entertaining representations, might (so far as his statement supposes) still lack the capacity to represent *laws*. And so one might justifiably think that Kant's view is that beings that cannot represent *laws* are not rational agents, that to have a will requires being able to represent laws.

Now one might think this because one supposes that the representation of anything requires the representation of laws. But this would be a mistake. Suppose, for instance, that *y* is a painting by an abstract expressionist. And suppose *x* is the sketchbook of an art student who is studying the painting by reproducing it in her sketchbook. *x* is thus a drawing of *y*. In being a drawing of *y*, *x* could be a representation of *y*. Yet *y* contains nothing in it that is lawfully related to anything else. Then *x* represents *y*, but, if it is faithful, neither *x* nor *y* contains any lawful relations within them. A representation thus need not contain anything in it that is lawfully related to anything else. One might assume that the possibility of representation itself requires some organizing principles within the thing represented. But why? There are paintings, for instance, that have no elements and *ipso facto* no organizing principles of those elements – for instance, a red canvas. And our imagined artist could represent those paintings, too.

There could be creatures that represented the world as if some sort of abstract expressionist created it – completely without any regular relations between any elements of the representation – but nevertheless moved about that world in a regular, law-like way through means that do not involve

representations of that world. If that is how things are for other agents, then perhaps only rational beings represent the world as operating according to laws. Then, again *a fortiori*, it would follow that only rational beings would have a capacity to act in accordance with such representations. However, this doesn't seem to be what Kant is getting at. It doesn't seem to be simply for lack of thoughts about the lawful operation of the world that non-rational agents lack a rational will. One could, for instance, imagine a creature who could represent the world as operating according to laws, in fact could represent itself as connected through laws to the operations of that world, and yet not have these representations lead to any action, to any willings, at all. Its representations of the lawfully related elements of the world might be causally isolated from that world.

Whether or not lawfulness is in fact a necessary condition of the possibility of being represented, Kant himself of course might have, for some reason or other, thought that it was. But, even if he did, it does not seem that the reason he held that only a rational being can act according to its representations of laws has to do with such a belief. It has instead to do with the fact that only such a being can *act* on the basis of representations, only such a being's causal power operates on the basis of representations, rather than simply on the laws of nature themselves. And this is in turn because only such a being possesses the rational faculty required, both in terms of representing and in terms of making things happen on the basis of those representations. In fact, this explains what Kant says next (*G IV* 412):

Since *reason* is required for the derivation of actions from laws, the will is nothing other than practical reason.

This statement should not be misunderstood of course. Kant is not here making the disastrous mistake some commentators have attributed to him, that since the will is practical reason, we cannot will irrationally, hence we cannot will immorally, hence we are not responsible for immorality.¹² We can indeed will immorally according to his view. In so doing, we are simply failing to conform our wills to principles that, insofar as we are fully rational agents, we will that we adopt and act on. Indeed, the will is a faculty that must employ reason to bring about action. Thus, only a rational being can act according to the representation of laws precisely because acting according to such representations requires reason to 'derive' an action from the law.

¹² See, e.g. Robert Paul Wolff's *The Autonomy of Reason* (New York: Harper and Row, 1973), p. 211.

Kant here says that reason is required to derive an action from a law itself, not the *representation* of a law. This is important. One might assume that Kant is saying something akin to what has puzzled Aristotle's readers for millennia, namely, that the conclusion of a practical syllogism is an *action*.¹³ That is, one might assume that Kant is representing practical reasoning in a way that is completely parallel to the way that many have thought that Aristotle represented it, with the major premise being the agent's representation of some law of action, the minor premise being the agent's representation of her particular situation (say, a perception) as covered by the law, with the action being the outcome of reason's putting these two things together. Reason's work does not appear to be of this sort, if we take Kant at his word in this passage. Reason derives actions from laws, where we are to understand that this is all conceived of as occurring within the realm of reasoning, rather than partially within that realm and partially in the realm of actions and events. The idea is that rational beings represent laws, laws represented as connecting their actions to the world through these laws, and in so doing they derive representations of their actions from these laws. Thus, practical reasoning occurs all within representations. Because it takes reasoning to derive anything from anything, it takes reason to derive an action from a law: in other words, to reason about how actions are lawfully related to the world.

For instance, given my aim to drink the water in the glass on the table and the belief that picking up the glass is a necessary means to drinking it, it takes reason to derive the representation 'Pick up the glass' from the representation of the law 'Take the necessary means to your ends'. Having derived this conclusion, however, I still must bring it about that I pick up the glass. Nevertheless, the correct explanation for my picking it up, should I do so, will have to include the fact that I formed the intention to pick it up by way of deriving it from this principle of instrumental rationality – a representation of a law.

Now consider in general how the representation of a law would be involved in the determination of the will. How, for instance, could a belief whose content concerns a law lead us to decide on a course of action? One obvious way in which it could do so is this: the law is a natural causal law and concerns how we are through our acts to cause some end that we have adopted. In that case, believing that a given causal law holds will lead us to act. Believing the law 'Touching fire causes burning sensations' will lead me, under circumstances in which I am aware of being in the presence of fire, to

¹³ Aristotle's *Nicomachean Ethics* (2nd edn, T. Irwin (trans.), Indianapolis/Cambridge: Hackett, 1999), 1147a27.

decide not to touch it, since *ceteris paribus* I am averse to burns. Believing the causal law 'Eating candy produces pleasure' will lead me, under circumstances in which I am aware of being in the presence of candy, to decide to eat the candy, since *ceteris paribus* I desire such pleasure. I say *ceteris paribus*, because of course there are other background conditions that must be met for me to decide any of these things.

So the representation of a law can determine the will when reason has derived an action from that law and there is a background end that is produced by or realized in that action as represented by the resulting derivation from that law. In the cases we have just considered, the laws may seem only to include psychophysical causal laws, laws connecting material events and sensations. Practical reason's job would, in this circumstance, seem to be purely cognitive and theoretical: reason simply locates and represents the relevant natural causal law connecting our actions to the fulfilment of our desires. However, Kant is not talking about the representation of *any* law as capable of determining the will, but clearly only those laws that are connected in the right way to the agent herself.

Return now to the (presumed) law of rational agency: 'Take the necessary means to your ends'. Much, if not all, of the time, 'take' and 'means' are causings and 'ends' are effects. Oversimplifying, the principle of the hypothetical imperative mostly amounts to 'insofar as you are a rational agent, cause the causes of your (aimed at) effects'. Since you aim to cause the effect, it is analytic that you cause the causes of those effects, by way of the principle noted above, that to conceive of x as the cause of y is to think that if there is some z that causes y , then either x causes z or $x=z$. However, even though 'take' and 'means' are both causal terms, in fact they are radically different kinds of causes. 'Means' can be thought of as causes that operate according to natural causal laws, but the idea of 'taking' those means is the idea of being a *first* cause, of being the origin of the causal chain referred to as the means that will result in an end. So, it turns out that one very surprising feature of Kant's conception of rational agency is this: *in every bit of rational agency*, we are acting according to our representations of *two* entirely different sorts of law, the laws governing the means, or phenomenal causation, and the law governing the exercise of our rational agency, a law of intelligible causation.

It is these sorts of considerations, along with the passages that support them, that seem to me to make a very strong case for the view that the moral law is a special sort of causal law. The moral law is the law according to which the causality of a rational will operates. And it must be universal because the very idea of a causal law requires universal validity.

3 PROBLEMS

One reason for thinking that Kant took the moral law to be a law of causality was that he discussed some of the difficulties this posed for his view. In the second *Critique*, he states that Hume was ‘quite correct in declaring the concept of cause to be deceptive and a false illusion’ given he ‘took objects of experience as things in themselves’ (*CpVV* 53). By contrast, he thought that since his transcendental idealism holds that the objects of experience are ‘only appearances’:

I was able not only to prove the objective reality of the concept of cause with respect to objects of experience but also to *deduce* it as an a priori concept because of the necessity of the connection that it brings with it

leaving him an embarrassing question, namely:

How is it with the application of this category of causality ... to things that are not objects of possible experience but lie beyond its boundaries?

That is, Kant’s line of argument, both in the *Groundwork* and in the second *Critique*, requires us to apply the concept of causation to rational beings. Indeed, ‘in the concept of a will ... the concept of causality is already contained, and thus in the concept of a pure will there is contained the concept of a causality with freedom’ (*CpVV* 55). But the exercise of free rational agency is not a possible object of experience, indeed *precisely because* it is free and hence not governed by natural causal laws. Since it is not a possible object of experience, there seems to be no justification for applying the concept of causation to any rational being.

Kant’s solution to this potentially devastating problem was to claim that our purpose in applying the concept of causation to rational agents is a *practical* not a *theoretical* one. I take this to mean that because we are reasoning our way to conclusions about what we ought to do, and not to conclusions about what we are, we are justified in applying the concept of causality to *noumena* such as rational wills. In so reasoning, given the fact that thoughts about ‘what we ought *to do*’ are at bottom thoughts about ‘what we ought to *cause*’, we have to make use of the concept of causation. We are ‘authorized’ to apply the concept in this case (*CpVV* 56):

by virtue of the pure, not empirical origin of the concept of cause, inasmuch as I consider myself authorized to make no other use of it than with regard to the moral law which determines its reality, that is, only a practical use.

That is, since the moral law presents itself to us as a law of causing an action whether there is any prior desire in us to do so or not, it is a law of causing an action whether there is any prior natural cause in us that will have this action as its effect. And since it is a law of causing an action whether there is any prior natural cause in us that will have this action as an upshot, it is a law of a free causation. This ‘fact of reason’ ‘determines the reality’ of our wills as free causes but only for practical purposes.

Thus, it seems very much as if Kant’s argument is that since he has shown that the concept of causation is justifiably applied to appearances, it now can be applied to things as they are in themselves *for practical purposes*. There are further problems that Kant does not discuss. For one thing, causation, insofar as it is a concept the analysis of which yields conformity to universal laws, is a *schematized*, not a pure, category. Unschematized, the concept is merely the logical connection ‘if-then’, logical antecedent and consequent, with no temporal dimension. It is its schematization that makes it possible for us to make phenomenal causal judgements that imply that events are lawfully related (*CrV* A146–7/B185–7). But if it were to apply to noumena, causation could not possibly be a schematized category, and so could not possibly yield the notion of conformity to universal law. Causal judgements about agents do not merely apply a concept that is only theoretically justified for use in the world of appearances; they apply a concept that *is itself* partly empirical. Indeed, it is just this fact that allows Kant to draw, by appeal to the concept of causation, the conclusion that, because rational agents are causes, they are bound by universal laws.

In Kant’s defence, however, the first *Critique* argues that we can at least think of causation in two ways, as *natural* causation or as *free* causation, and it is the latter ‘*transcendental* idea of *freedom* on which the practical concept of freedom is grounded’. The faculty of free causation in us is (*CrV* A533/B561):

the faculty of beginning a state *from itself*, the causality of which does not in turn stand under another cause determining it in time in accordance with the law of nature ... a spontaneity, which could start to act from itself.

What is needed is some way of getting from this idea of spontaneity, of a being’s ‘beginning a state from itself’ to the idea of conformity to universal law, *without* appealing to the concept of ‘the connection of a state with a preceding one in the world of sense upon which that state follows according to a rule’, that is, without appealing to the concept of natural causation. The problem is that free causality does not, as he puts it, ‘stand under any conditions of sensibility and is not itself appearance’ (*CrV* A 539/B 567).

Kant appears to think he has gained the notion of conformity to universal law in this way: when we think about the world governed by natural causation and hence universal natural laws, we are thinking about everything that has happened, is happening, or will happen. But when we consider the idea that reason might be a spontaneous cause of something in the world, 'we find a rule and order that is entirely other than the natural order', the order, not of what happens, but what *ought* to happen (*CrVA* 550/B 578). Since it is still an order of 'happenings' it is still a causal order. The 'non-natural' laws of free causation govern this alternative 'non-natural' causal order. While this thought goes some way toward helping us to understand why and how Kant thought that causation can be rightfully deployed in practical matters, it still leaves unexplained why this non-natural free causation should be ordered lawfully, given it seems that it is the concept of natural causation that carries with it the idea of universally valid laws.

Perhaps there is some further idea that will render Kant's views internally consistent on this point, though I fail to see what it is. A final issue nevertheless remains. Any defence of the formula of universal law will have to rely on a conception of causation according to which a causal relation between particular events implies that there is some universally valid law governing the types of events that are so related. But it is by no means uncontroversial that the concept of causation implies such law-governed relations. There could be counterfactual accounts, for instance, that comport well with the idea that there is some sort of metaphysical connection or 'glue' between causally related events, just the sort of metaphysical glue that regularity theories deny. And while the existence of laws may well in turn be required to explain these connections, whether it does or not depends, not upon the analysis of the concept of causation, but the metaphysical structure the universe. What this means for Kantians who would like to defend the universal law of nature formulation of the categorical imperative is that, if I am right, they will not only have to plug the gap between conformity to universal law and the canonical statement of that formulation; they will also have to defend the idea that causation implies the existence of universal laws governing the causal relata.

4 THE ROLE OF REASON IN RATIONAL AGENCY

If conformity to universal laws is a requirement of agency rather than rationality, what is left for the 'rational' in rational agency to do? I want to end on a few thoughts about this. What is distinctive about the laws of

rational as opposed to other sorts of law-governed causality? As I see it, Kant's thought was that reason brings spontaneity, not universality, to the moral law. Whatever the universally valid law is according to which the will is a cause, it is the law of a spontaneous 'first' cause. In part, I believe he thought this because of the way in which reason operates, or at least as we understand or sense how reason operates (*CpVV* 48, V 101). When reasoning, either theoretically or practically, we do so with some sense of our reason as operating with a kind of spontaneity. This spontaneity, moreover, is not apparent when, for instance, we observe people or (if possible) ourselves reasoning. Hence, Kant states that (*CrVA* 547/B 575):

in the case of lifeless nature and nature having merely animal life, we find no ground for thinking of any faculty which is other than sensibly conditioned. Yet the human being, who is otherwise acquainted with the whole of nature solely through sense, knows [*erkennt*] himself also through pure apperception, and indeed in actions and inner determinations which cannot be accounted at all among impressions of sense.

Kant thought that we somehow 'know' through 'pure apperception' that we are spontaneous origins of actions. I put 'know' in scare quotes because counting this as knowledge runs counter to Kant's official line that knowledge extends no further than the objects of possible experience, and he is here denying that we can know about this spontaneity through some sort of empirical introspection. That this apperception is 'pure' implies that it is not an empirical awareness of oneself. So Kant believes we are in some special sense conscious of ourselves and how our reasoning operates, but not through empirical introspection. Kant adds 'and indeed in actions'. It is not merely the spontaneity of reason in theoretical reasoning, in reasoning from belief to belief, of which we are (somehow) cognizant; we are also aware of the spontaneity of reason in practical reasoning, in reasoning from imperatives to action.

Apperception is supposed to be some kind of consciousness of oneself of the sort contained in connecting 'I think' to all of our representations. Although I myself am unable to make completely clear what it is that Kant is thinking of with the idea of 'pure apperception', I have some idea of what he is getting at. With regard to our experiences of the world, it is difficult to imagine that someone could be unsure of whose experiences he was having. Of course, it is trivially true that my experiences are mine. And some schizophrenics feel that their thoughts aren't really theirs or that they are not the authors of their thoughts, that their thoughts in fact belong to someone else who is beaming them into their consciousness. But in some

sense beyond the bare analytic truth, yet which is consistent with psychotic phenomena, we are aware that our experiences are necessarily ours. Likewise, it is difficult to imagine someone who could be unsure of whose volitions he was having. Again, in some sense beyond the related bare analytic truism, but consistent with the phenomena associated with some kinds of psychoses, we are aware that our volitions are necessarily ours.

Now, turning to the awareness of the spontaneity of our reason: what Kant himself says (again in *Groundwork* III) is this (*G* IV 448):

In [a rational agent] we think of a reason that is practical, that is, has causality with respect to its objects. Now, one cannot possibly think of a reason that would consciously receive direction from any other quarter with respect to its judgements, since the subject would then attribute the determination of his judgement not to his reason but to an impulse. Reason must regard itself as the author of its principles independently of alien influences.

The idea expressed in this passage, I take it, is roughly this: it is impossible to imagine someone who could be unsure whether his reason operated spontaneously through a given bit of reasoning that he engaged in. A person is apparently necessarily aware of the spontaneity of the operation of his reasoning. There is some plausibility to this: consider reasoning through a maths or a logic problem. It is not that we have an *experience* or *feeling* of our own reason spontaneously moving from thought to thought, nor do we *believe* our reasoning faculty's movements from premise to conclusion or through steps of addition is spontaneous. Indeed, we may believe quite the reverse, that this sense of our reason as operating spontaneously is an illusion. Nevertheless, in such activities, we do seem to be operating under some conception of our own reason as not 'made to' reach conclusions from outside of itself, as moving freely from premise to conclusion. Reason is making itself move through premise to conclusion, from figures to sums, and so on. That, at any rate, seems to be a datum, if not of experience, then of an overall conception under which we reason.¹⁴

Likewise, we have a sense of our reason's spontaneity provided by apperception of the operation of practical reason as a cause. Given Kant's own arguments regarding natural necessity, no natural laws govern this causality. Yet there can be no causality without laws. So our apperception of the spontaneity of reason is left with a vacant place where a law should stand. 'Pure practical reason', Kant boldly states in the second *Critique*, 'fills

¹⁴ Thanks to my colleague Alexander von Schönborn for pointing out this passage. His helpful paper covers Kant's views on intelligible causation: 'Kant and the Absolute', *Southwestern Journal of Philosophy* 7 (1976), 145–52.

this vacant place with a determinate law of causality in an intelligible world (with freedom), namely the moral law' (*CpV V 49*). If our reason is a spontaneous cause, or at least if we have to operate under that conception of ourselves, then the law according to which our causality as agents operates is the law of a spontaneous causality. And, of course, the law of a spontaneous cause can only be a law of which that cause is itself the author. So whatever else reason brings to rational agency, it brings the 'auto' part of the idea of autonomy, the idea that this universal law according to which it must operate originates in the cause itself.