



The Wrongdoing that Gets Results (2004)

by John Gardner
Professor of Jurisprudence
University of Oxford
<http://users.ox.ac.uk/~lawf0081>

This is an author eprint, which may not incorporate final edits.
The definitive version of the paper is published in

Philosophical Perspectives 18 (2004), 53
doi: [10.1111/j.1520-8583.2004.00021.x](https://doi.org/10.1111/j.1520-8583.2004.00021.x)
© John Wiley and Sons 2004

The typescript appears here with the consent of the publisher,
under the publisher's eprint policy, or by author's reserved rights.
Please do not quote from or cite to this eprint. Always use the
definitive version for quotation and citation.

The Wrongdoing that Gets Results

JOHN GARDNER *

1. *Reasons to try and reasons to succeed*

Some reasons for action are reasons for trying, and some are reasons for succeeding. I have reasons, just at this moment, to write a good paper. I also have reasons to try to write a good paper. It is tempting to suppose that reasons to try and reasons to succeed always go hand-in-hand like this.¹ Surely I have reasons to try to write a good paper if and only if I have reasons to write a good paper? Surely, indeed, my reasons for writing a good paper just *are* my reasons for trying to write one? In which case, the suggestion that I have reasons to try and reasons to succeed, while not false, is misleading. It gives the impression that, just at this moment, I have a double ration of reasons for action. I have reasons for trying, and on top of that reasons for succeeding.

* Professor of Jurisprudence, University of Oxford. The first version of this paper was written nearly ten years ago, during my tenure of a British Academy research leave award. A version closer to the present one was presented in 1998, at the Analytical Legal Philosophy Conference in San Diego. Belated thanks to all the participants for their interventions, and especially to Michael Moore, who acted as discussant. The paper has made various appearances in various guises since, but never before in print. Among the many valuable points that were made at seminars and conferences, comments by Jo Wolff, Martha Klein, and Sabina Lovibond have proved particularly helpful.

¹ And correspondingly reasons *not* to try and reasons *not* to succeed. I should just make clear, for the avoidance of doubt, that throughout this paper I treat 'not trying' under the 'trying' head and 'not succeeding' under the 'succeeding' head. Thus, as some of my examples will indicate, the argument of the paper applies, *mutatis mutandis*, to acts and omissions alike.

When in fact all I really have is one set of reasons, which are reasons to try to write a good paper and, by trying, to succeed.

It is true that, as in this case, reasons for trying are normally reasons for succeeding, and vice versa. But it should be noted that this relationship of identity can be structured in two quite different ways. Reasons for succeeding are *integrally* reasons for trying when they are reasons for succeeding-by-trying. One does not do as these reasons would have one do if one tries without success, but nor does one do as they would have one do if one succeeds without trying. In fact, in many cases involving reasons to succeed-by-trying, success is partly constituted by trying, so that, strictly speaking, success without trying is no success at all.² There is no reason to play a game of squash or football without trying to. Indeed there is no such thing as playing a game of squash or football without trying to. These activities are partly constituted by the attempt to engage in them. We have reasons in these and many other cases to be successful in the attempt, but not otherwise. There is much of interest to be said about such reasons. For the moment, however, let me just draw attention to a fallacious line of argument which leads some to suppose that reasons for succeeding are *always* integrally reasons for trying.

All reasons for action, the argument goes, must be such that it is possible in principle for people to act on them. If one acts on a reason to succeed, then necessarily one tries to succeed. Thus, it is concluded, all reasons for succeeding are necessarily reasons to succeed-by-trying. The mistake here does not lie in the premiss that all reasons must be such that it is possible in principle for

² Sometimes the word 'success' is used exclusively to denote success in pursuit of one's goals (as in 'she's a very successful banker'). In this usage the word implies at least intentional success if not success-by-trying. Success in this intentional sense is of great moral importance because successful pursuit of one's (worthwhile) goals lies at the heart of one's well-being. In this paper, however, I speak of 'success' without this implication of intention because I mean to raise the prior question of what should *be* one's goals. Should one aim only to try? Is trying the most that is rationally required of one?

people to act on them. Reasons for succeeding meet this condition (as the subsequent premiss concedes) and as long as the words 'in principle' are taken seriously it is a sound condition. The idea that there are reasons for action on which no rational being could ever act is indeed impossible to stomach, and here we find a grain of truth in the idea that 'ought' implies 'can'. So the mistake in the argument does not lie in the first premiss. Nor does it lie in the second premiss, with which one might certainly quibble, but only regarding some borderline cases the classification of which is not relevant to the issue now under discussion. The mistake lies, rather, in the hidden third premiss that, since reasons must be such that they can in principle be acted on, all that matters, rationally, is which reasons we act on. Of course, it is true that it matters, rationally, which reasons we act on. It matters rationally not only because some reasons for succeeding are integrally reasons for trying, but also in various other ways. But what is important for now is that which reasons we act on is not the *only* thing that matters rationally. What is rationally important, first and foremost, is that we do whatever the reasons that apply to us would have us do. Many reasons for action are such that it is possible in principle for us to do precisely what they would have us do, whether or not we do so by acting on them. That can be true of reasons to try and reasons to succeed alike. There are reasons to try such that we do exactly what they would have us do if we try, whether or not we try for those reasons. And there are reasons for succeeding such that we do exactly what they would have us do so long as we succeed, whether or not we succeed by acting for those reasons, and, more broadly, whether or not we succeed by trying to.

This brings us to the second class of cases in which reasons for succeeding are also reasons for trying. In these cases reasons for succeeding are *derivatively*, rather than integrally, reasons for trying. Suppose that I telephone my mother on her birthday, when I inadvertently press the wrong memory button on my telephone in the course of trying to telephone somebody else. In

that case I made the telephone call I had reason to make without trying to make it. No doubt some of the reasons I had for my telephoning my mother on her birthday – e.g. that making such a call expresses my love for her – are reasons to succeed-by-trying, so that I do not do as they would have me do when I telephone her in this fortuitous way, without trying to. But other reasons – e.g. that she is expecting me to call – are reasons such that I do exactly what they would have me do if I telephone her, even if I do it without trying to. As long as these latter reasons to telephone my mother on her birthday still apply (that is, until I have telephoned her, or her birthday comes to end, or she finally gives up waiting and telephones me) they are also reasons to *try* to telephone her. Once these reasons to telephone no longer apply, I no longer have the relevant reasons to try. But for as long as my trying is unsuccessful (the line is busy, or my mother has gone out shopping) I have reason to try again until I succeed, or until it emerges that no amount of trying will do the trick (my mother's telephone is out of order, or my mother has gone camping in the mountains). I have these derivative reasons to try telephoning my mother just in order to succeed in telephoning her, and they continue to apply to me for as long as trying will contribute to the success I have reason to achieve.

The word 'contribute' here is deliberately ambiguous. It raises the question of *how*, i.e. by what logic, reasons to try derive from reasons to succeed in the cases in which reasons to succeed are derivatively reasons to try. The answer, which I have already built into my explanation of the telephoning case, is that they derive by what Anthony Kenny calls 'the logic of satisfactoriness'.³ According to the logic of satisfactoriness I have reason to do whatever is sufficient to achieve whatever I have reason to achieve, whether or not it is also necessary. It should be emphasised that the logic of satisfactoriness does not require that

³ Kenny, 'Practical Inference', *Analysis* 23 (1966), 65. The analysis is usefully sharpened and reinforced in Kenny, *Will, Freedom and Power* (Oxford 1975), 81ff.

trying be *logically* sufficient for success in order to derive reasons to try from reasons to succeed. After all, trying to telephone my mother never strictly *entails* that I will succeed. For in order to constitute success in this enterprise, my trying to telephone must be combined with the call being answered at the other end, which is, relative to my trying, a contingency. Nevertheless, in the sense required, trying to telephone can be sufficient for success. It is sufficient for success just in case, by trying, I will at last succeed. In that case my reason to succeed is also a reason to try. The application of this logic explains why, once no amount of trying to telephone my mother can bring success, I will no longer have derivative reasons to try, and yet, in the meantime, I continue to have derivative reasons to try even though I could equally telephone my mother without trying to, by pressing the wrong memory button. My reasons for succeeding are also automatically reasons for me to try if and only if circumstances are such that trying will in the end bring success. It matters not whether, in the same circumstances, success might equally come to me without my trying.

Now the application to reasons of this logic of satisfactoriness means that there are counterexamples to our initial hypothesis that reasons for succeeding and reasons for trying always go hand-in hand. There are, at any rate, cases in which I have a reason to succeed but no reason to try. Suppose that, since I cannot swim a stroke (and have no boat, and no helicopter, and no telephone, and am perched on a cliff-top in the middle of nowhere, etc.) it would be quite futile for me to try to rescue a man who is drowning in the stormy sea below. That this man needs to be rescued is a reason for me to rescue him. If I had no reason to rescue him, after all, I would not be so horrified at the realisation that it would be so utterly futile for me to try. I could walk past without compunction. But, by the logic of satisfactoriness, the futility of my trying does have the consequence that my reason to save the man is not a reason for me to try to save him. No amount of trying on my part will

allow me to save him. Here we can see one of the many false attractions of the claim that 'ought' implies 'can'. When the word 'ought' is given its advisory (as opposed to mandatory) inflexion, this slogan taken literally asserts that I have no reason to succeed whenever it is the case that I cannot succeed. In fact, however, it is not the existence of a reason to succeed, but rather the implication of a reason to try *from* a reason to succeed, which is blocked by the impossibility of success. It is the common assumption that reasons to try and reasons to succeed always go hand-in-hand that leads many to suppose that if, in a given case, there are no reasons to try in order to succeed, then equally there can be no reasons to succeed, so that 'ought to succeed' does indeed imply 'can succeed'. My horror, as I look down from the edge of the cliff on the drowning man below, bears witness to the falsity of this view. The horrible truth is that I have a reason to save this man's life even though I cannot save it, and therefore, other things being equal, have no reason to try.

Some may venture an alternative interpretation of my horror. They picture me wavering on the cliff's edge, incapacitated by indecision, now leaning forward to jump, now pulling back. Doesn't this suggest the impetus of a reason to attempt a rescue? If so, doesn't that point away from the view that reasons adhere to the logic of satisfactoriness, and towards the opposite view that, sufficient or no, the mere fact that trying to effect a rescue is necessary to effect a rescue creates a reason to try out of my reason to succeed? In that case, the futility of trying is presumably significant only as a countervailing consideration, pulling me back even as my reason to try pushes me forward. My horror is, on this view, the horror of the classic moral dilemma. But this explanation of what is going on cannot be right. The main reason *against* my attempting a rescue, in the drowning man case, is obvious. I will end my own life by hurling myself into a sea from which, as a non-swimmer, I cannot hope to escape. If the attempt at rescuing the drowning man in the process were not futile, this could conceivably expose me to a dilemma: Preserve

my own life or take a chance on saving his? But the fact that an attempt to rescue the drowning man is futile eradicates rather than complicating this particular dilemma. It does not add an extra reason against making an attempt, but removes the relevant reason in favour, since it leaves me with no chance of saving anybody's life. It extinguishes one horn of the dilemma. So if I am indeed wavering on the edge of the cliff, pondering whether to jump, then the explanation for my doing so must be different. One possible explanation is epistemic, based on my uncertainty. I am wondering whether I am, contrary to fact, capable of effecting a rescue, and so wondering whether my reason to succeed may, after all, be a reason for me to try. Another possible explanation is that I am a man who is, in momentary defiance of logic, trying to succeed without trying to, because he knows that he has a reason to succeed, but, other things being equal, no reason to try. Whichever of these explanations we go for, the truth remains that I cannot regard my reason to effect a rescue as a reason to try unless I detect a possibility of success. Far from pointing away from the logic of satisfactoriness, then, the sight of me wavering on the cliff's edge, incapacitated by indecision, helps to confirm its application, complementing rather than rivalling what I already said about the true horror of my situation.

It may be thought that there is a third possible explanation for my wavering on the clifftop that is still compatible with the logic of satisfactoriness. Maybe this is a case in which I have an independent reason to try, not derivative of any reason to succeed. Mightn't I have a reason, for example, to express my undying love for the drowning man in an act of futile self-sacrifice? Perhaps this is what is impelling me to jump even though a successful rescue is admittedly out of the question? Certainly I may have such independent reasons to try. There are reasons to try that are not reasons to succeed just as there are reasons to succeed that are not reasons to try. And yet one may well doubt whether such independent reasons to try could serve

to explain my wavering on the cliff-top. Someone who recognises that his trying will not contribute to his succeeding cannot conceivably try. This is because of the nature of trying. Trying is acting for the reason that one's action will (supposedly) contribute to one's succeeding. If I misjudge my prospects of success – if it seems to me that maybe I am capable of saving the drowning man after all – then I am capable of trying because I am capable of supposing that my trying will contribute to my succeeding, and *ex hypothesi* I have a reason to succeed. Of course, I am wrong. My only possible real reason for trying, in such a case, is one that is independent of my reason to succeed (e.g. a reason to express my undying love for the drowning man in an act of futile self-sacrifice). But this cannot conceivably be the only reason for trying that I *take* myself to have when I try (or waver on the point of trying). I must also take myself to have a derivative reason for trying, a reason to succeed that is also, derivatively, a reason to try. So a non-derivative reason to try, of a purely expressive kind, can rationally reinforce a derivative reason to try, and make it rational, in appropriate cases, to try disproportionately to the contribution that such trying will make to success. What such a non-derivative reason to try cannot do is make it rational to try when one knows that, from the point of view of success, there is nothing to be gained by trying.

2. Duties to succeed: introducing Kant's argument

To recap: There are reasons to succeed-by-trying (the integral case), and reasons to try just in order to succeed (the derivative case). There are also reasons to succeed which are not reasons to try, and reasons to try which are not reasons to succeed. But can all the same things be said of *duties* to try and *duties* to succeed? Since duties are reasons – or to be more precise, the fact that one has a duty is a reason⁴ – you might assume so. But many doubt it.

⁴ To save words, I will use the less precise formulation throughout.

Considering my cliff-top predicament, many people are happy to accept that I have a *reason* to save the drowning man, while denying that I can conceivably have a *duty* to do so. If they are right, the explanation must lie in one or other of the two features which combine to distinguish duties from ordinary reasons.

A duty is a *categorical mandatory* reason for doing what one has a duty to do. What is mandatoriness? As Joseph Raz has explained, the mandatory force of a reason comes of the fact that it is 'protected' by second-order reasons against acting on some or all of the countervailing reasons.⁵ In everyday language, this protection makes it the case that one is *bound* or *required* to do what the reason would have one do. Again care should be taken not to read into this the assumption that, whenever a reason is protected in this way, one is bound or required to act *on* the protected reason. Nor should one be tempted into the even more elaborate assumption that, whenever a reason is protected and one is accordingly bound or required to act, one is thereby bound or required to act on the consideration that one is bound or required to act. Possibly there are cases in which one has not only a protected reason to act as one does but also a protected reason to act as one does for that or some other protected reason. But the ordinary or default case of a protected, or mandatory, reason is free from these trappings. Like non-mandatory reasons, mandatory reasons are in the ordinary case just reasons for one to do what they would have one do, irrespective of why one does it. Their mandatoriness resides in the effect they have on the rational acceptability of acting on certain countervailing considerations, and therefore doing the opposite of what the mandatory reason would have one do, but it does not, in the ordinary case, affect the rational acceptability of acting on other reasons which point towards the same action that the mandatory reason itself requires or binds one to perform.

⁵ Raz, *Practical Reason and Norms* (London 1975), 73ff, read to take account of the important modification in Raz, *The Authority of Law* (Oxford 1979), 18.

The *categorical* quality of a reason resides, meanwhile, in the fact that it is not hostage to the prevailing personal goals of those to whom it applies. I have a reason to go shopping now, if I want to eat at home tonight. That is a non-categorical (a.k.a. 'hypothetical') reason, because it applies only on condition that I still want to eat at home tonight. I have a reason to go shopping tomorrow, on the other hand, because I already invited people to dinner tomorrow night. That is a categorical reason because it continues to apply to me irrespective of whether I still want to entertain my guests when the time comes to go shopping. And it is a duty if it is also mandatory, i.e. if I am bound to do what it would have me do by the presence of second-order reasons not to act on at least some of the countervailing reasons, such as reasons of convenience or expense. You may object that its being categorical automatically entails that it will be mandatory, since when a reason is categorical the reason 'I want to do otherwise' is a countervailing consideration which is by definition excluded from consideration. But that is simply not true. When a reason is categorical, the reason exists and applies to me irrespective of what I want at the time. Whether what I want at the time can still count in its own right as a reason which weighs *against* the categorical reason is another matter altogether. My own view is that what I want at a given time is never, at that time, a reason in its own right for or against my doing anything, and so can never weigh in the balance against any reason, categorical or otherwise. I will return to this point below. But even if what I want can be a reason for me at the time when I want it, it is one thing to say that some of the other reasons which apply to me apply to me irrespective of what I want at that time (i.e. they are categorical) and quite a different thing to say that what I want at that time is not to be weighed in the balance against those other reasons (i.e. they are mandatory). Being mandatory is one property all duties share, and being categorical is another. There are mandatory reasons which are not categorical, and categorical reasons which are not mandatory, but reasons in these two classes are not duties.

One familiar view, which I will not explore here, is that it is the mandatoriness of duties which, in the example of my cliff-top predicament, prevents me from having a duty to save the drowning man even though I have a reason to save him. The claim that 'ought' implies 'can' is admittedly unsound when 'ought' is given its regular advisory inflexion, the story goes, but we should stand by it when 'ought' is given its mandatory inflexion. This would rule out the possibility of a duty to succeed wherever, as in the drowning man scenario, there is no possibility of success. The view that the mandatory 'ought' implies 'can' is frequently associated with Kant.⁶ But Kant does not endorse this view. In fact, as we will see, he repudiates it, carefully avoiding each and every one of the fallacies which have lent it its spurious credibility. Nevertheless, Kant certainly concurs with adherents of the view that the mandatory 'ought' implies 'can' in holding that, in my cliff-top predicament, I have no duty to save the drowning man. Kant's line of argument emphasises the fact that duties are categorical rather than the fact that they are mandatory. And far from attempting to show that I can have no duty to succeed when I cannot succeed, his argument purports to show that (barring certain special cases) I can have no duty to succeed *irrespective* of whether I can succeed. In fact his argument purports to show that (barring certain special cases) I can have no duty to succeed *full stop*. It thus has implications for the other examples we discussed in above, apart from the drowning man case. For instance, it may be that I have a duty to try telephoning my mother on her birthday, and indeed I may have every reason to succeed when I try, but, according to Kant's argument, quite irrespective of the possibility of success I can have no duty to succeed in telephoning her. This is the radical argument that I will be considering here.

⁶ Most famously by David Ross in *The Right and the Good* (Oxford 1930), 4–5.

On the most plausible interpretation of Kant's argument, the possibility or impossibility of success is irrelevant *both* to the existence of duties to try *and* to the non-existence of duties to succeed. On this interpretation his conclusion has two limbs. Limb one: It may be true of some non-categorical reasons for trying that I do not have these reasons when I cannot succeed, but, since duties are categorical reasons, this cannot be true of duties to try. Limb two: it may be true of some non-categorical reasons for succeeding that I can have these reasons even when I cannot succeed, but, since (except in the special class of cases to be introduced shortly) there is no such thing as a categorical reason to succeed, and duties are categorical reasons, this in turn cannot be true of duties to succeed. I will for the most part ignore the first limb of this conclusion, and restrict my attention to Kant's argument insofar as it supports the second limb. Pared down to relevant essentials, the argument goes like this:

(a) The only source of unconditional value in our actions is the good will;

(b) the good will infects not the whole of what we do but only that part of it which consists in our trying to do good;

(c) doing one's duty is of unconditional value;

thus (d) there can be no duty to succeed, but only a duty to try (unless trying entails success).⁷

Let me just elaborate a couple of points which are perhaps made a little opaque by the skeletal form of the argument. First, the parenthetical proviso at the end of the conclusion (d) is needed to cover the possibility that, in some cases, there is no logical gap between trying and succeeding. If someone asks me to try my

⁷ *Groundwork of the Metaphysic of Morals* (ed. H.J. Paton, London 1948), 62 ('the good will and its results') combined with 67–8 ('the formal principle of duty').

hardest then I succeed in doing as asked if and only if I try my hardest. There is, logically speaking, no more to this particular success than my trying. Kant's argument does not rule out the possibility of my having a duty to succeed in a task that consists only of endeavour. But it does rule out the possibility of my having a duty to succeed in the endeavour itself, even if the endeavour is easy and I am sure to succeed. That is because the latter is an empirical, and not a logical, inevitability. It is conditional upon the state of the world beyond my will, and thus, for Kant, cannot affect the incidence of unconditional value in my action as specified in premiss (a). The parenthetical proviso covers only logical inevitabilities, cases where trying involves succeeding *by definition*, and so requires no conditions to be met in the world beyond my will: hence the word 'entails'. In this special case, but not otherwise, Kant allows that one may have a duty to succeed.

Second, premiss (c) is deliberately expressed nonexclusively. It does not say that doing one's duty is the only thing that is of unconditional value. One might think, on first inspection of the argument, that the point in premiss (c) could be extended to reasons in general. Isn't there, after all, unconditional value in doing what reasons would have one do, namely the value of rationality? Recognising this extension would transform Kant's argument into an argument against the possibility of any reasons to succeed, subject always to the parenthetical proviso in (d), and that would controvert almost all of what I said in section 1 above. But Kant did not endorse such an extension. This is where the distinction between categorical and hypothetical reasons comes to the fore in his thinking. It is the fact that duties are categorical reasons that makes their performance, in Kant's view, unconditionally valuable. The value of action in accordance with hypothetical reasons is, according to Kant, conditional. Its value is conditional on the value of the prevailing personal goal on which the application of the reason depends. Not so categorical reasons. They are reasons which are dictated by the good will

rather than by prevailing personal goals, and the value of action in accordance with them is therefore, by premiss (a), unconditional. To the validity of these last moves we will need to turn our attention in due course. For the time being, they suffice to explain why Kant's argument is supposed to cut against the existence of duties to succeed but not against the existence of all reasons to succeed. The existence of hypothetical reasons to succeed is not affected by the argument, since action in accordance with them is not unconditionally – or, as Kant also put it, *morally* – valuable.⁸

3. *The limits of Kant's argument*

In simple terms, the argument just introduced is supposed by Kant to trace the moral rectitude of our morally right actions to what we put into them as opposed to what comes out of them. This has been remembered by many as the 'Kantian' view that moral agency is a luck-free zone: that good luck cannot embellish, and bad luck cannot tarnish, our moral agency. Bernard Williams speaks of the 'Kantian conception' of morality which insists 'not only [on] morality's immunity to luck, but [on the agent's] own partial immunity to luck through morality.'⁹ Thomas Nagel agrees, characterising the very same argument of Kant's which I set out above as an argument to the effect 'that good or bad luck should influence neither our moral judgment of a person and his actions, nor his moral assessment of himself'.¹⁰ But is this really the burden of Kant's argument? Is what

⁸ *ibid*, 84.

⁹ Williams, 'Moral Luck', in his collection *Moral Luck* (Cambridge 1981), 20 at 21.

¹⁰ Nagel, 'Moral Luck', in his collection *Mortal Questions* (Cambridge 1979), 24 at 24. Nagel couches much of his subsequent discussion in the language of 'responsibility' rather than that of 'wrongdoing' or 'duty'. I will disregard this distracting feature here and assume that he does mean to persist with the same philosophical problem as Kant and Williams.

Williams calls the ‘Kantian conception’ really Kant’s own conception? Let’s put on one side for a moment – in fact until near the end of this paper – the obvious difficulty that the word ‘moral’ and its cognates have fluctuating and sometimes competing nuances depending on context, so that we may wonder whether Kant and his commentators are always *ad idem* in identifying the subject-matter of their inquiry. Let’s assume with Kant, for the time being, that the word ‘unconditional’ in premisses (a) and (c) of his argument is univocal and interchangeable with the word ‘moral’. Is the argument then compatible with, let alone supportive of, the view that there can be no such thing as moral luck? That is not my reading of it. I regard the argument as much more limited, and believe it was regarded by Kant himself as much more limited, than is now customarily supposed. Indeed I think Kant’s argument militates, in some respects, against the more sweeping ‘Kantian’ view with which it has since become associated. Although there are several others, I will mention here just two key disparities.

3.1 Doing one’s duty and violating it. Contemplating Kant’s argument, which proceeds from explicit consideration only of the value stemming from a *good* will, Nagel writes: ‘He would presumably have said the same about a bad will: whether it accomplishes its evil purposes is morally irrelevant.’¹¹ But this certainly cannot be presumed. Kant’s argument in fact yields an important asymmetry between the doing of one’s duty and the violating of it. I expressed Kant’s premiss (c), as Kant himself often expressed it, in a rather misleading way. Kant did not exactly hold that every action of doing one’s duty is an action of unconditional, or moral, value. On the contrary, he accepted that we may do our duty out of motives other than that of the good will, and that in such cases, although our actions lack the unconditional, or moral, value bestowed by his premiss (a), what

¹¹ *ibid*, 24..

we do nevertheless *still counts as doing our duty*. Kant was right to accept this. While there are cases in which, if we do not do our duty for the right reasons, we do not do our duty at all, these cases constitute the exception rather than the rule. In the normal run of cases, doing one's duty is still doing one's duty irrespective of one's reasons for doing it. Even someone of ill will, who does her duty only to spite her detractors, at least has it to be said in her favour that she does her duty. Nasty though she may be, she is no wrongdoer. Kant agrees. We know that Kant agrees because he believes, and stresses throughout his moral philosophy, that duties can be enforced by ordinary earthly legal systems.¹² Legal regulation can be used, in other words, to compel people of ill-will to do their duty. If being of ill-will were incompatible with doing one's duty then this legal enforcement would not be a logical possibility.

To accommodate this point, we need to put a slightly more elaborate gloss on his premiss (c). We should perhaps recast it: Every duty is such that its performance can be of unconditional value. Read with premiss (a), this means (or should be interpreted to mean) that every duty is such that acting out of a good will is logically sufficient, albeit not logically necessary, to perform it. Those of ill-will may do their duty; but those of good will *must*, by definition, do their duty. This rendition of premiss (c), I will concede for the sake of argument,¹³ preserves the argument to the conclusion (d) that there can be no duties to succeed, except where trying entails success. Except where trying entails success, obviously, acting out of a good will is not logically sufficient to perform a duty to succeed. There always remains the logical possibility of failure, even with the best will in the world. Even with the best will in the world, one's actions may not turn out the way they are meant to. Thus, on this rendition of premiss (c), there can still be no duty which is done

¹² *The Metaphysics of Morals* (ed. Mary Gregor, Cambridge 1996), 21.

¹³ Arguably the refined (c), when made tolerably precise, is made intolerably conclusory. I waive the point to avoid a long digression of limited interest.

only on condition that one's action turns out in the right way – there can be no duty to succeed, unless trying entails success.

Yet this rendition of premiss (c) nevertheless quite clearly leaves open the possibility of a duty which is *violated* only on condition that one's action turns out in a certain way. Take, for example, the legal duty, violation of which constitutes the tort of negligence at common law. To comply with this duty, it is logically sufficient that one takes reasonable care, irrespective of how one's action turns out. There is a respectable interpretation of the law, however, according to which one's failure to take reasonable care does not amount to a violation of the relevant duty unless it also results in damage of a legally recognised kind.¹⁴ A duty with this structure passes Kant's test under premiss (c), since acting from a good (in this case, careful) will is logically sufficient to perform it. And yet its violation is partly constituted by the way it turns out, i.e. by the damage that is done. Thus the so-called 'Kantian' view that luck can neither embellish nor tarnish our moral agency is decidedly more symmetrical than the view to which Kant himself was committed. If we are of good will, for Kant, our moral agency is invulnerable to luck – that our actions turn out for the worse cannot transform our actions from right to wrong. We do the morally right thing *come what may*. Therein lies the unconditional value of acting from a good will. But if we are not of good will, luck in the way our actions turn out can contribute constitutively to our breach of duty. It can be part and parcel of our moral wrongdoing, tarnishing our moral agency. Some will doubtless read an implication of tarnished moral *character* into this talk of tarnished moral agency. But that is a mistake. I already explained why in the previous paragraph. Kant correctly thought that one could be a morally deficient person without invariably performing morally deficient actions.

¹⁴ The alternative respectable view is that the duty is violated whenever reasonable care is not taken, irrespective of damage, but that legally recognised damage is a requirement of *locus standi* for the violation to be actionable. Which view of the law is correct is of no concern to us here.

Even a person of ill will may do her duty. He did not accept, on the other hand, that one could perform morally deficient actions without being a morally deficient person, since, in his view, a person of good will necessarily does her duty. Once we know this much, it should be no surprise to learn that Kant tolerated the possibility that what contributes to the moral deficiency of one's actions need not be something that also contributes to one's status as a morally deficient person. The fact that a person of ill will does not do her duty need not be attributable solely to the fact that she is a person of ill will. The moral deficiency in her action, i.e. its being a breach of duty, could be constituted partly by the deficiency in her character, her ill will, and partly by the way that action turned out. The duty of care in the law of negligence, interpreted as above, works exactly like this. The wrongdoing involved in breaching this duty is only partly a matter of the wrongdoer's carelessness, a deficiency that may be traced to her character. It is also partly a matter of the damage she does by her carelessness. Kant's argument must be read as tolerating this possibility, and hence, by premiss (c), as tolerating the possibility of moral luck in the evaluation of our actions to at least this limited extent.

3.2 How our actions turn out and how we turn out. The example of the duty, violation of which constitutes the tort of negligence, also helps to bring out a second limitation on Kant's argument as an argument against the possibility of moral luck. The duty of care in the English law of negligence is a duty to take reasonable care, not a duty to take as much care as one can. It is what lawyers call an 'objective' standard. If the duty of care is conceived as a moral duty – and remember that, by Kant's premiss (c), all duties must be conceivable as moral duties – then its objective standard exposes us to an element of moral luck, viz. luck in whether we, as we now are, can live up to that standard. The standard is more demanding, and less easily attainable, for some people than for others. For some people, or on some

occasions, it may even be unattainable. Although attainable with the best will in the world, some people plainly do not have the best will in the world. Their will is weak, or corrupted by sloppiness or selfishness or poor judgment. In English law, that is tough luck: one's liability in the law of negligence is unaffected. But could it also be tough *moral* luck? Or does morality, faced with people who do not have the best will in the world, reduce its demands to a level at which they can nevertheless manage to do all that morality would have them do? Kant unequivocally rejects the latter possibility in favour of the former. 'We must not determine moral duties,' he writes, 'according to our estimate of a man's power to fulfil the moral law; on the contrary we must estimate man's moral power by the standard of the moral law, which commands categorically. Hence we must appraise this power on the basis of our rational knowledge of what men should be ... not on the basis of our empirical knowledge of men as they are.'¹⁵ In this and other passages Kant commits himself to what I earlier called the 'grain of truth' in the claim that 'ought' implies 'can'. He agrees that the demands of duty are demands which can, in principle, be met. People, as rational beings, can aspire to the qualities of character which duty demands, can aspire to have the best will in the world. But Kant roundly rejects the suggestion that we can always, *in practice*, live up to the standards of duty. He rejects the suggestion that when we have an imperfect moral character, i.e. less than the best will in the world, the demands of duty come down to meet us. What is more, he regards himself as compelled to reject that suggestion by the logic of his own argument. Only the good will is a source of unconditional value. Every duty must be such that doing it can be an act of unconditional value. The demandingness of our duties must therefore be dictated exclusively by the standard of the good will, not by the standard of the will that we happen to

¹⁵ *The Metaphysics of Morals*, above note 12, 164 (my translation rather than Gregor's).

have. The standard of duty is thus, by Kant's own reasoning, an objective standard in the lawyer's sense. Whether we do our duty cannot, for Kant, be hostage to luck in the way our actions turn out; but the corollary is that it can and must be subject to luck in how we turn out as agents, in whether we have the personal qualities needed to live up to duty's demands at the time when those demands are made of us. In Kant's morality, as I put it before, the moral rectitude of our actions depends on what we put into our actions rather than what comes out of them.

Some may raise their eyebrows at my equation of the idea in the last sentence with the idea in the sentence preceding it. For some agents, as I already conceded, the very same standard may be more demanding and less easily attainable than for others with different personal qualities. So if the standard of duty is constant, or objective in the lawyer's sense, then some people need to make more effort in order to do their duty than other people do. Surely, in that case, a person's moral rectitude doesn't after all depend on what she puts into her actions, but depends on something else instead? This question does much to reveal where the misunderstanding of Kant lies, that has transformed Kant's view of morality into the view of some of today's self-styled 'Kantians', in which a person's moral rectitude is relativised to her existing moral competence. Kant's reaction to this view can only be that, however hard he may try, the person of lower moral competence does not put as much into his actions, morally speaking, as does the person of higher moral competence. The morally relevant input, for Kant, is not effort but good will. Lacking an equally excellent will, the person of lower moral competence necessarily puts less moral value, even as he puts more effort, into whatever he does. Charged with (say) lying or cheating, such a person may say: 'I always try as hard as I can to tell the truth, in fact I tire myself out with trying, but my instinct for self-preservation or self-advancement always gets the better of me.' Kant's answer to that is: 'Given your admittedly imperfect moral character, your weakness of will, there are cases in which

trying as hard as you can is not trying hard enough to do your duty.’ Our underachiever may respond (with the blessing of some ‘Kantians’): ‘My character was a product of my environment, so how can it be morally significant?’ To which Kant himself answers: ‘I didn’t say your character was morally significant. The whole point is that, for the purposes of assessing morality’s demands, your character is entirely *insignificant*. Morality demands what it demands unconditionally, and that means, among other things, irrespective of your character. For a start, it makes demands of your character irrespective of your character: Whatever you are like, and however you came to be like that, you should be a person of good will. On top of that, and more crucially, morality makes demands of your actions irrespective of your character: Whether or not you are a person of good will, you should still do your duty just as if you were a person of good will. To those who complain that, since they are not people of good will, there are times when they do not and (given their personal qualities) cannot do their duty, what can I, Kant, reply but ‘tough luck’? How often must it be said that the only way to guarantee that your life will not be not blighted by moral wrongdoing is to be a person of good will.’

On this point, Nagel again appears to misread Kant. ‘To Kant,’ he says ‘it must be possible for everyone to achieve [moral virtue] by making the right choices, against whatever temperamental background.’¹⁶ This threatens to turn Kant’s view on its head. For Kant, it is true that everyone, whatever their other personal qualities, could become a person of good will. But it is also true that, given their other personal qualities, many people are not people of good will, and can only become people of good will if they lose some of those other personal qualities. The latter group of people are under constant peril of committing moral wrongs. Their actions, unlike the actions of people of good will, are not guaranteed automatically to meet

¹⁶ Nagel, ‘Moral Luck’, above note 10, at 33.

morality's demands. Whether or not they do meet those demands may be, relative to them, a matter of good or bad moral luck in how they are constituted as people. Maybe they never enjoyed an adequate moral education, or had relationships in childhood that left them with weak characters, bad tempers, or addictive personalities. Kant's argument does not make such moral luck inconceivable. On the contrary, it makes it inevitable. It follows that the proviso at the end of Kant's conclusion (d) – 'unless trying entails success' – covers rather more cases than at first some might have imagined when I labelled such cases 'special'. There are plenty of cases in which the only reason why people's attempts are not logically sufficient for success is that the hardest they could try, given their personal qualities, was not hard enough. Some people try as hard as they can, given their personal qualities, to resist temptations, but still they fail. Kant's argument does not rule out duties to succeed in such cases. It does not rule out the possibility, for example, of a duty not to lie. Lying is an action which by its nature can only be committed intentionally. Telling an untruth without meaning to is not lying. Thus one may avoid lying altogether by never forming an intention to tell an untruth, and forming an intention is a matter which falls, in principle, within the domain of the will. It is no answer for those who are accused of violating this duty to say that they tried and tried not to lie but still they failed, so trying obviously did not entail success for them. It is enough, for the purposes of Kant's argument, if trying entails success for people of good will. If there are others for whom trying does not entail success then all that follows, so far as Kant's argument is concerned, is that they are not people of good will. Perhaps they are thereby made victims of moral luck, but it is not a kind of moral luck against the possibility of which Kant's argument is directed, nor against the possibility of which Kant's argument can intelligibly be deployed. Kant's argument is directed and deployable only against the possibility of moral luck in the way that our actions turn out – and even there, as we saw, only

against the possibility of good luck being needed for moral rectitude, not against the possibility of bad luck being needed for moral turpitude. Kant's position on moral luck is accordingly a long way from what Williams and Nagel represent as the 'Kantian' position. Kant's morality is not a luck-free zone, and our moral agency is not, for Kant, comprehensively immune to luck. In fact, to put it concisely, our moral agency's immunity to luck under Kant's morality arises only on the strict condition that we always do, whether by luck or otherwise, what people of perfectly good will would do. Then and only then do we necessarily enjoy, as the person of good will does, the utter moral irrelevance of our action's results.

4. The failure of Kant's argument

Kant's conclusion, then, does not rule out the possibility of actions being morally wrong in virtue of their results, so long as the wrongdoing in question can only be committed through a personal failing or lapse such as carelessness, dilatoriness, or laziness. But it does rule out the possibility of actions which are morally wrong in virtue of their results with no such threshold fault condition – what lawyers sometimes call 'strict liability' wrongdoing. Nagel mistakenly collapses these two possibilities in explaining Kant's moral position. He observes that in cases of negligence people are held 'responsible for the contributions of fate as well as for their own – provided they have made some contribution to begin with', commenting that, according to Kant's approach, this seems 'akin to strict liability, which may have its legal uses but seems irrational as a moral position.'¹⁷ But when we shift attention from Kant's 'approach' to Kant's argument, the difference between a result-sensitive negligence liability and a result-sensitive strict liability turns out to be absolutely pivotal. The best will in the world guarantees, in logic

¹⁷ *ibid*, 31.

and not merely in fact, that one does not incur the former liability, whereas it cannot ensure, in fact let alone in logic, that one will not incur the latter. It is against moral instantiations of result-sensitive 'strict liability' wrongdoing, accordingly, that Kant's argument is exclusively directed.

Directed, yes, but successful? I will argue not. My argument has two stages. The first represents a challenge to the plausibility of Kant's premiss (a). The second contends that Kant's anticipatory answer to this challenge, which is indeed the only possible answer I can think of, is either itself based on confusion or else turns the argument to conclusion (d) into a striking non-sequitur. Either way, the argument to (d) fails.

4.1 The doctrine of moral autonomy. As already mentioned, Kant holds – and rightly so – that acting for non-moral reasons is compatible with doing the morally right thing. The implication of acting for non-moral reasons is not that one breached one's duty but merely that one's action of doing one's duty lacked genuinely moral value. So the legal enforcement of duties, as I said before, is no contradiction in terms. Ultimately, indeed, moral recalcitrants can be brought round to doing their duty by the use of a variety of incentives which appeal to them non-morally: threats of punishment, offers of reciprocation, promises of personal gain, etc. The problem is that this process of adding non-moral reasons for action cannot possibly endow the actions of moral recalcitrants with any moral value. That is because the moral value of an action is closely related to the way in which it manifests (and, in return, is constitutive of) the moral character of the person whose action it is. Other things being equal, honest people need only the facts which make an action honest to motivate them to perform that action, trustworthy people similarly regard the facts which make an action trustworthy as sufficient reason to perform it, fair-minded people need only the facts which make the action fair, etc. Accordingly, those who need extra reasons before they are persuaded to perform the same

actions simply show themselves to be less than honest, short on trustworthiness, lacking in fair-mindedness, etc. That is why such people's actions cannot be made more morally valuable by providing such people with extra, non-moral reasons to do the morally right thing. Their moral deficiency, which impoverishes the moral value of their actions, lies in the very fact that they *need* such extra reasons, that they are not already motivated to do the morally right thing all by themselves. Noticing this important truth leads Kant to describe moral value as a kind of 'inner value'. It is value which has its source within us, stemming from our moral character. Moral reasons, correspondingly, are for Kant self-legislated reasons: they are reasons that we give ourselves, and which cannot, without losing their moral quality, emanate from anywhere else. This is Kant's doctrine of moral autonomy.

It does not follow from Kant's doctrine of moral autonomy alone that premiss (a) holds, i.e. that the only source of moral value in our actions is the good will. That follows only if, in addition, the good will exhausts our moral character. Needless to say, Kant believes that it does. He holds that the good will is all that is left when one subtracts from the character of a rational agent all those traits which give rational salience to the agent's own prevailing personal goals, and which therefore contribute to the hypothetical (non-moral) rather than categorical (moral) dimension of reasoning. In fact Kant goes further and denies that the value which these latter traits of character lend to our actions meets the autonomy condition in the first place – all reasons conditional upon prevailing personal goals are, in Kant's view, directed towards external incentives and acting on these reasons endows our actions, accordingly, with no 'inner value'. For Kant, in other words, there is no such thing as genuinely autonomous reasoning which is not moral reasoning. But be that as it may, what matters now is the converse claim which Kant also makes. There is, for Kant, no such thing as moral reasoning which is not autonomous reasoning. That is to say, all the moral value in our actions has its source within us. And the only source

of moral value within us is the good will, that part of one's character which has no dealings with one's prevailing personal goals. Hence premiss (a).

The main difficulty in this line of thought lies in the very idea that moral value has a source. It is true that *legal* reasons and values are source-based, i.e. are valid in law because someone or something made them so. Kant's idea of morality as a kind of personal legal system – 'the moral law within us' – famously carries this feature over into morality. But reliance upon it nevertheless seems to put Kant's doctrine of moral autonomy at odds with the very features of moral virtue upon which it depends for its intuitive appeal. In the eyes of the honest person, as I said before, the facts which make an action honest are, all else being equal, reason enough to perform it. For the trustworthy person the same holds regarding the facts which make an action trustworthy. And so on. Those who manifest these virtues, in other words, look not upon their virtues but upon facts about their actions as their reasons for performing those actions. The honest person alights on the fact that he is replying to a morally acceptable question, and the fact that a certain reply would be (to the best of his knowledge) true, as adding up to sufficient reason, *ceteris paribus*, for making that reply. The trustworthy person regards the fact that a certain action was promised by her, and that its performance according to the promise has now fallen due, as sufficient reason, *ceteris paribus*, for performing that action. And so on. Morally virtuous people, accordingly, do not regard the value in their actions as stemming from their virtues, but as stemming, if stemming is the right word, from their actions. And in fact stemming does not seem to be the right word at all. As Aristotle saw, the value in their actions is simply, to morally virtuous people, the value in their actions, so it need not move from any *source* to any *destination*, e.g. from their character to their actions. It lies where it lies, and is reason enough without

needing any further source of value.¹⁸ Thus moral reasoning is not, in the eyes of the morally virtuous, source-based at all. If this is a correct analysis of moral virtue, and assuming that moral virtue involves no rational delusion, then Kant's doctrine of moral autonomy rests upon a mistake. Kant jumps from the sound proposition that only morally right actions which manifest moral virtues have moral value to the conclusion that moral virtue must be the source of moral value in morally right actions. He leaps over the much more natural explanation, consonant with the logic of moral virtue as analysed above, that the moral value in our actions lies in our actions, but is realised *qua* moral only on condition that we see it there and act on it. Moral virtue, in other words, is not a source of moral value in our actions but merely a *condition* of moral value in our actions. This modified doctrine of moral autonomy preserves the important truth that the very fact of needing extra incentives to perform morally right actions betokens moral deficiency, and hence leaves one's actions *pro tanto* morally impoverished. It is therefore true to the valuable insight underlying Kant's version of the doctrine. But at the same time it eradicates the spurious legalisation of morality which Kant's version of the doctrine imports: it eradicates the idea that moral reasons are source-based reasons. And by the same token it invalidates premiss (a) of Kant's argument against the possibility of duties to succeed, since (a) presupposes that moral value is source-based. This remains so even if we continue to assume, with Kant, that the good will exhausts our moral character. Even on that assumption, the modified doctrine of moral autonomy entails only the modified version of premiss (a):

(a') There is unconditional (moral) value in our actions only on condition that we perform them from a good will.

Substituting (a') for (a) means that Kant's argument against the possibility of simple 'strict liability' duties to succeed never gets

¹⁸ Aristotle, *Nicomachean Ethics*, 1105^a31.

started. That is because (a') has no consequences for how much of what we do is infected by moral value, and hence does not speak to (b); (a') sets a threshold for moral value in our actions but has no implications for its extent or reach. That being so, Kant's argument is doomed from the start.

4.2 Unconditional virtues. The above remarks stand or fall on the view that the intuitive appeal of the doctrine of moral autonomy can be traced to the features of moral virtue that I highlighted. The only possible answer that I can think of – and Kant's own answer – is that I have the nature of moral virtue quite wrong. Either the thinking of honest, trustworthy and fair-minded people is not as I described it, or else honesty, trustworthiness and fair-mindedness are not truly moral virtues. Kant makes use of both of these alternatives in anticipating my kind of challenge. Some virtues commonly thought of as moral, such as courage and loyalty, are exiled to the realm of non-moral; other virtues commonly thought to be as I described them, such as honesty and trustworthiness, are reanalysed so that their moral flavour is preserved.¹⁹ The pivotal point in all this is that moral virtues, on Kant's analysis, require more *self-consciousness* than I allowed. To be truly honest, it is not enough to regard facts such as the fact that one is making a true response to an acceptable question as a reason for one's action. One must see these facts and act on them *under a moral description*. One must think of making true responses to acceptable questions as imposing a duty, and act on them *for the sake of duty itself*. Otherwise any virtue one manifests is a non-moral virtue. Thus one must indeed, as a morally virtuous person, have a source in mind for the value in one's action. That source is the good will, the legislator of duty. So (again assuming that moral virtue comports no rational illusion) premiss (a) stands.

One possible 'Kantian' ground for thinking that moral virtue is self-conscious in this way can be quickly disposed of. It comes

¹⁹ *Groundwork*, above note 7, e.g. at 61, 66.

of the thought that those who do not act for moral reasons under a moral description are in fact acting, instead, on their prevailing personal goals. That being so, surely their reasons for acting become hypothetical reasons rather than categorical ones, and their virtues correspondingly become, on Kant's view of morality, non-moral virtues? There is a simple confusion here. When an honest person does what she does for the reason that doing it counts as giving an true answer to an acceptable question, she may well *want* to tell the truth and moreover tell the truth *because* she wants to. She may want to tell the truth precisely because, as an honest person, she recognises the value in giving a true answer to an acceptable question. It does not follow that she acts for a hypothetical reason, or even that she sees herself as acting for a hypothetical reason. It may be the case that she should respond as she does even if she were, contrary to fact, *not* inclined to do so, and that, being honest, she knows this to be the case. My own view, an essentially Aristotelian view, is that the impeccably honest person is always inclined to tell the truth in response to a morally acceptable question, because reasons of honesty permeate not only her actions but also her personal goals. In this dimension she wants exactly what she should want, and wants to do exactly what she should do. Indeed I subscribe to an Aristotelian version of the doctrine of moral autonomy according to which the moral value in our actions depends (with some notable exceptions) on our prevailing personal goals being aligned with morality's demands, so that we are in any case inclined to do as morality would have us do.²⁰ That is precisely why, to my mind, morally virtuous people need no extra incentives to do the right thing. Kant sets himself against this Aristotelian view of moral autonomy.²¹ But whatever cause he

²⁰ *Nicomachean Ethics*, 1104^b4ff.

²¹ How radically is a matter of dispute among commentators. Some say that for Kant there is neither moral loss nor moral gain in having one's inclinations aligned with morality's demands, so long as one is inclined to act for the sake of duty. Others paint Kant's doctrine in a more austere light: inclination not

may have for doing so it cannot be that the Aristotelian view of moral autonomy and moral virtue is incompatible with his view that moral reasons are by their nature categorical reasons. There is no incompatibility between these views at all. Reasons are categorical just in case they apply to us irrespective of our prevailing personal goals at the time, i.e. whatever we want. It does not follow that acting *for* such reasons is valuable irrespective of our prevailing personal goals. As Kant himself said, and I have reaffirmed throughout this paper, it is one thing to do as moral reasons would have one do, to do the morally right thing, and another to do it for moral reasons, and hence with the added benefit of moral value. We should do our duty even if we do not want to – duties are, after all, categorical reasons – but it does not follow that it would not be morally better to do our duty spontaneously, wanting to do it. Recall that according to the view I earlier expressed wanting to do something is not, in itself, a reason to do it. That one does something because one wants to is therefore not anathema to one's doing it for moral reasons, even though moral reasons are categorical reasons. On the contrary, it seems to me that one does something for moral reasons whenever (i) one does it because one wants to and (ii) one wants to do it for moral reasons. So if Kant were to rely, in support of his doctrine of moral autonomy and hence in support of (a), on the thought that those whose virtue is not self-conscious are necessarily acting on hypothetical and therefore non-moral reasons, his thinking would be seriously confused.

Although some 'Kantians' may follow this confused line of thought towards premiss (a), and hence ultimately towards conclusion (d), Kant himself seems not to do so. Leaving aside the occasional incautious remark, he does not trace the self-

only does not aid but positively poisons moral reasoning. I have taken pains in my discussion not to allow either of these interpretations, which both have some textual support, to gain ascendancy. The best discussion I know is Barbara Herman, 'On the Value of Acting from the Motive of Duty', in her volume *The Practice of Moral Judgment* (Cambridge, Mass. 1993).

consciousness of moral virtue directly to the fact that all moral reasons are categorical reasons. But he does directly trace it, all the same, to the fact that all moral reasons are unconditional reasons. It turns out that the unconditionality of moral reasons does not stop, according to Kant's account, at their categorical nature. Moral reasons apply to us irrespective of our prevailing personal goals, to be sure; but they also apply to us, in Kant's view, irrespective of who we are, where and when we live, what the consequences of following them will be, what other people do, what other people want, what traits of character we have, and so on. Indeed moral reasons, as Kant conceives them, are conditional only upon themselves; the good will is a purely self-referential source of reason and value. It is none other and no more than the will which respects the will which respects the will which respects the will ... and so on *ad infinitum*. The question of whether this self-referential 'universalisation' test for the validity of moral reasons is empty has attracted a great deal of attention.²² For myself I would go further and question whether the idea of a reason conditional only upon itself is intelligible. But let us leave these doubts on one side for the moment. The point which concerns us here is that Kant's adherence to the self-conscious view of moral virtue is closely related to the universalisation test for moral validity. That is because, according to the universalisation test for moral validity, the good will has no other concern but the complete self-consistency of the good will. The focus of attention in genuinely moral reasoning must therefore ultimately be turned inwards on the quality of the will to be invested in an action. That being so, moral virtues cannot be as I explained them above. Recall my explanation of the virtue of honesty: The honest person treats the fact that such-and-such an answer would be (to the best of his knowledge) a true answer to an acceptable question as sufficient reason, *ceteris*

²² An excellent survey of the debate is included in Christine Korsgaard, 'Kant's Formula of Universal Law' in her *Creating the Kingdom of Ends* (Cambridge 1996).

paribus, for giving that answer. This is not, according to Kant's universalisation test, a moral reason for action. That is not because by the universalisation test there can be no moral reason to give a true answer to an acceptable question. On the contrary, Kant could quite easily regard giving (what is to the best of one's knowledge) a true answer to an acceptable question as a matter of duty: for this is clearly a duty in which trying entails success, and indeed a very short step from the duty not to lie. Rather the point is that by looking on the action itself as if it were the source of the moral reason for performing it one introduces an alien whiff of conditionality into moral reasoning. One does not see the action in its unconditional guise, as an act of respect for the moral law of the universalised will. In this unconditional guise the reason for giving such-and-such an answer is not the mere fact that such-and-such is a true answer to an acceptable question. The reason is the fact that such-and-such is a true answer to an acceptable question *coupled with the fact that giving a true answer to an acceptable question is one's duty*. No reasoning is sufficiently unconditional to count as moral reasoning, and no will is accordingly a good will, without this ultimate attention to its own moral standing, its own fidelity to the moral law. Without this attention the moral agent is like a lawyer who happens to rely upon the correct legal doctrine in her argument but has completely overlooked the legal authority from which that doctrine takes its validity. The doctrine is then correctly relied upon, but not correctly relied upon *qua* legal. The lawyer is to that extent exposed as a deficient lawyer. Likewise the moral agent who is motivated to give a true answer to an acceptable question without regarding this as an act of respect for the will which respects the will which respects the will ... and so on. She alights on the morally correct course of action but has completely overlooked the authority, the authority of the universalised will, from which it takes its moral validity. She is not motivated to do the morally right thing *qua* moral, *qua* duty. Her will is not a

good one, her virtue is accordingly not a moral one, and her actions are not invested with moral value.

Let us suppose that this manoeuvre is not only intelligible but plausible. If so, it certainly supports premiss (a) of Kant's argument over the rival (a') that I adduced above. But it does so only at the high price of rendering the argument from (a) to (d) invalid in a new way. For it turns out that the doctrine of moral autonomy upon which premiss (a) depends for its appeal can be defended only by reinterpreting the word 'unconditional' in premiss (a) to mean something much more demanding than merely 'categorical'. It now means not merely 'irrespective of the agent's prevailing personal goals', but rather 'irrespective of anything at all apart from itself'. Yet there is no reason at all to read this more demanding sense of 'unconditional' into the independent premiss (c) of Kant's argument. There is no reason – and Kant himself offers no reason – to suppose that every duty need be such that its performance can be unconditionally valuable in this more demanding sense. And so the argument to (d) becomes a clear non-sequitur. There can be duties which are such that the strictures of premisses (a) and (b) do not apply to them, since their performance is not unconditionally valuable in the premiss (a) sense; thus there can, so far as Kant's argument is concerned, be straightforward duties to succeed irrespective of any element of trying – in other words, straightforward 'strict liability' wrongs.

The point is quite easily illustrated. If I invited people to dinner tonight and I have no food in the house, then as I already mentioned that gives me a categorical reason to go shopping for food this afternoon. I have a reason to go shopping irrespective of my prevailing personal goals, i.e. whether or not I still want to entertain. But this reason to go shopping is not a reason that applies to me irrespective of the consequences of going shopping. If it is Christmas Day today and every food shop that I could conceivably get to is closed, then other things being equal I have no reason, categorical or otherwise, to go shopping today. That

is because going shopping cannot possibly have the consequence of enabling me to entertain my friends to dinner tonight. The logic of satisfactoriness applies no less to categorical reasons than to other reasons: other things being equal, I have no categorical reason to do what is insufficient to achieve what I have categorical reason to achieve, even though it may be necessary. So there can certainly be categorical reasons which are also hostage to consequences. These reasons are, of course, excluded from the realm of the unconditional, or moral, by Kant's defence of premiss (a) which is based on his doctrine of moral autonomy and his associated universalisability test for moral validity. These reasons depend for their application on something other than the good will itself.²³ Yet in reality they may well be duties. Unless we deny such things the status of duty by arbitrary stipulation, it is perfectly conceivable that I have a duty to go shopping for food this afternoon just so long as that will have the consequence that I can thereby do my further duty in entertaining my friends tonight. In fact many duties are of just this consequence-sensitive type. If the word 'unconditional' in premiss (c) is given the same demanding meaning that it has in premiss (a), so as to preserve the logic of the argument, then these consequence-sensitive duties straightforwardly falsify premiss (c): for the value in performing such duties is never consequence-independent, and therefore never unconditional in the premiss (a) sense. If, on the other hand, 'unconditional' in premiss (c) is still read as 'categorical', so that these examples do not falsify the premiss, then the argument to (d) is invalid. There can be duties which escape the strictures of premisses (a) and (b) and Kant's argument therefore gives us no reason to doubt that these duties could be straightforward duties to succeed. Kant's argument therefore gives us no reason to interpret my duty to go shopping for food this afternoon as a duty to try to go shopping, or more broadly as

²³ Hence Kant's adherence to the hitherto suppressed 'limb one' of his argument (see section 2 above), by which he effectively denies the application of the logic of satisfactoriness within morality.

a duty which is not breached so long as I tried hard enough to go shopping. Even with the best will in the world, that I will do the shopping when I try to cannot be guaranteed. Perhaps I will be hit by a car as I cross the road to the supermarket, or wrongfully arrested at the checkout. Even then, even with this raw bad luck potentially bearing on how my attempt to do the shopping might turn out, my not doing the shopping for tonight's dinner may be a straightforward breach of my duty, a 'strict liability' wrong. Kant's argument, even if we grant his implausible doctrine of moral autonomy, does nothing to eliminate this possibility. For it has, *pace* Kant, no implications for the structure of consequence-sensitive duties. Thus Kant's argument fails.

Straightforward though this shopping example is, it also helps us to see where confusions may creep in which help to lend undeserved credibility to Kant's position. Let me mention four of these confusions, which have been to greater or lesser extents influential in giving succour to the so-called 'Kantian' view of morality to which Nagel and Williams refer. First, the very notion of a 'goal', which figures in the contrast between categorical and hypothetical reasons, may mislead one into thinking that there cannot be such things as consequence-sensitive duties. People have goals in the sense that there are things that they want. Actions may also have goals in the sense that there may be consequences which their agent intends to bring about by performing them. One may easily slip into confusing people's goals with the goals of their actions. Thus, it may seem, a reason which applies independently of prevailing personal goals, i.e. categorically, must also be consequence-insensitive. But of course people may have the personal goal of performing certain actions – for example, keeping promises or telling the truth – irrespective of the consequences, i.e. without those actions having any goal. And conversely, an action may have a goal which is not, or no longer, a personal goal of its agent – for example, going shopping for food may have the goal of enabling me to entertain friends I have already invited to dinner

tonight even though right now I have no inclination at all to be sociable. People's personal goals and the goals of their actions are not the same things at all. Thus reasons which apply to me irrespective of my prevailing personal goals are not by the same token reasons which apply to me irrespective of the consequences of my doing as they would have me do, nor *vice versa*. That duties apply to us irrespective of our prevailing personal goals does not entail, accordingly, that duties apply to us whatever the consequences. Kant himself sometimes courts this confusion, which contributes significantly to the now widespread confusion about the exact relationship between deontological moral views and consequentialist ones. The foregoing remarks make clear that there can in principle be moral views which emphasise the importance of duty above all, and which in fact include no reasons for action other than categorical ones, and yet which harbour no special insensitivity to consequences. But a confusion between people's goals and the goals of their actions may lead one to think otherwise.

Secondly, it is tempting to suppose that even if the categorical nature of duties cannot make them consequence-insensitive, the mandatoriness of duties can. But this supposition repeats a mistake that we already encountered earlier when explaining the difference between the categorical and the mandatory. That duties are mandatory means that they are protected reasons, reasons for an action coupled with second-order protective reasons which exclude at least some countervailing reasons from consideration. Among those reasons which are excluded from consideration may of course be found, in some cases, countervailing reasons which are based on the adverse consequences of the mandated action. But this has nothing to do with the question of whether the protected reason *in favour* of the mandated action depends on the positive consequences of doing as that reason would have one do. It may be true both (i) that I have a duty to go shopping for food this afternoon if and because that will enable me to feed my friends

tonight, and (ii) that the fact that this is my duty means that I have reason to disregard some of the bad consequences of going shopping for food such as the fact that it will make me overdrawn at the bank and the fact that I will be very tired when my friends arrive. Element (ii) certainly makes my reason protected and hence mandatory, but it does not by the same token make it consequence-insensitive, because (i) still holds. So again consequence-sensitive duties remain a real possibility. That remains so, indeed, even with so-called *absolute* duties, which exclude from consideration *all* countervailing considerations. Since the consequences I am speaking of when I speak of consequence-sensitive duties do not countervail against the duty but on the contrary support it, they are unaffected by the absolute exclusion. Even absolute duties may therefore be consequence-sensitive duties. Again today's widespread confusion about the relationship between deontological and consequentialist moral views sometimes reflects, or is compounded by, the mistaken thought that this cannot be so. Deontological views are sometimes vaguely associated with an absolute mandatoriness which is thought to make them anti-consequentialist. But in fact the absoluteness of a duty has no implications at all for its consequence-sensitivity, and hence for its intelligibility to consequentialists.

Thirdly, one may be tempted to suppose that contrary to the conclusion I drew from the example, the very fact that I have no duty to go shopping for food to entertain my guests if it is Christmas Day and the shops are shut shows that, after all, I can have no duty to go shopping, but at most a duty to try. This is another variation on a mistake we encountered earlier. If my going shopping for food cannot enable me to entertain my guests, then other things being equal my duty to entertain my guests yields *neither* a duty to go shopping *nor* a duty to try to go shopping. That is a straightforward application of the logic of

satisfactoriness.²⁴ It has no bearing on the proposal that there can be no duty to succeed but at most a duty to try, because it is a case in which there is neither. Or rather there is neither so far as the shopping is concerned. *Ex hypothesi* there still remains, of course, the prior duty to entertain my guests, which, given that the shops are shut, I must now fulfil in some other way, e.g. by taking them out to dinner. If that too is impossible then sadly it looks as though I am doomed to breach my duty to entertain my guests. It will be no answer to say that I tried to shop for them and tried to take them out but, with the best will in the world, failed in both ventures. For these attempts did not exhaust my duty. My duty was to entertain my guests to dinner, not to try (ever so hard) to do so. Thus the example which I used to discredit Kant's argument certainly does not lend accidental support, *malgré lui*, to Kant's conclusion (d).

Finally, it may be thought that my argument was viciously circular. I invoked a case of a consequence-sensitive duty to assist

²⁴ I should just mention here, to forestall objections, some complications with the application of the logic of satisfactoriness to duties and other mandatory reasons. That I have a mandatory reason to achieve something entails that I have a derivative reason to do whatever will be sufficient to achieve it. But the latter reason need not itself be mandatory. The derivative reason will be mandatory only if it mandates some action which is *both* necessary *and* sufficient to achieve what I have mandatory reason to achieve. Say, for example, I have a duty to provide my children with a way of getting home from school. This yields a reason for me to meet them with the car, since that will be one way for them to get home. But it does not follow that I have a duty to meet them with the car. That follows only if doing so is the only way to get them home, i.e. if other ways for them to get home have been ruled out. Other things being equal I can fulfil my duty to give them a way home by giving them bus money or buying them bicycles or moving house to live next to the school. Until the possibilities are narrowed down, I have no duty to do any one of these things, but while my duty to provide the children with a way home remains I have a reason, of non-mandatory force, to do each of them. None of this casts doubt on the application of the logic of satisfactoriness to mandatory reasons. It only adds an additional requirement to that logic in a certain narrow class of cases.

in making out the claim that, *pace* Kant, there can be duties to succeed which are not exhausted by trying. But surely there can only be consequence-sensitive duties if there can indeed be duties to succeed, i.e. wrongdoing dependent on results? So surely I assumed my own conclusion? This objection is doubly confused. In the first place, it confuses the results of our actions with their consequences. My action of killing my enemy has the consequence that I am safe from further attack by him. But it does not have my enemy's death as a consequence. His death is rather a *constituent* of my killing him: if he has not yet died, then I have not yet finished killing him. The death is the *result* of the killing, which is to say that my action of killing my enemy is constituted not only by the death of my enemy but also by the causal contribution I made to his death. A result, to put it simply, is a causal constituent of an action, whereas a consequence is a causal subsequent of an action.²⁵ Thus my enemy's death is a result, not a consequence, of my killing him. True, my enemy's death is a consequence, rather than a result, of my *trying* to kill him. Logically speaking I have not finished killing him until he dies, but by then I have, logically speaking, finished trying to kill him. Typically, those who cast my enemy's death as a consequence rather than a result of what I did are assuming that the action of mine which matters in such a case is the action of trying to kill, rather than the action of killing. But in that case they are the ones who are trapped in the vicious circle. They are still assuming Kant's conclusion, which is the very conclusion that my argument undermines. They are assuming that I cannot have a duty to succeed, but at most a duty to try. A duty not to kill is a straightforward duty to succeed, i.e. not to bring about the death irrespective of what one tried to do. But it need not be consequence-sensitive. I may have such a duty irrespective of the consequences of my having it and performing it, e.g. irrespective

²⁵ In using the terminology of 'results' and 'consequences' to mark this distinction I have borrowed, like many others, from G.H. von Wright, *Norm and Action* (London 1963), 40-41.

of whether it has the consequence that I am not safe from further attack. So it is not viciously circular to base my argument for the existence of duties to succeed upon the existence of consequence-sensitive duties. In the second place, the accusation of vicious circularity confuses the grounds of our duties with the content of our duties. A consequence-sensitive duty is one that arises only on condition that the action it gives one a duty to perform will contribute causally to something else happening or being done. But it need not be a duty *to* contribute causally to anything. It may be a duty to keep a promise or get up early or go shopping for food, or indeed a duty to try and get a job or to try and start a family etc. It may be either a duty to succeed or a duty to try, in other words, but when it is the former it need not be a duty to succeed in bringing about the further events or actions, the pursuit of which it exists to serve. I may have a duty to keep a promise because if I fail to do so I will never be trusted again. But my duty is a duty to keep my promise, not a duty to make sure I am still trusted. It is a consequence-sensitive duty to succeed, but not a duty to succeed in bringing about the consequence to which it is sensitive.

A propos this last point I should add that for the purposes of my argument in this sub-section nothing substantial turned on the fact that I chose an example of a *consequence-sensitive* duty to show that there are examples which are excluded from the domain of the unconditional for the purposes of Kant's premiss (a) but are included in the domain of the unconditional for the purposes of his premiss (c). I could equally have used an example of a duty which is conditional upon time or place, or upon the actions or wishes of others, or upon the identity of the agent etc. For as I said, Kant's defence of his doctrine of moral autonomy, and hence his defence of premiss (a), depends on the claim that moral considerations are conditional upon *nothing but themselves*. Thus any example of a duty with an element of conditionality would do to make the key move in my argument. The key move is that Kant's argument becomes a dramatic non-sequitur as soon

as we see that ‘unconditional’ in premiss (c) means merely ‘categorical’, while ‘unconditional’ in premiss (a) means ‘categorical *and* consequence-insensitive *and* irrespective of time and place *and* independent of the actions of others and so on.’

5. *The pseudo-problem of moral luck*

With the failure of Kant’s argument, what becomes of the so-called ‘Kantian’ denial of the possibility of moral luck? Since I know of no other argument against the possibility of moral luck apart from the one which Kant himself makes, and with which I have been dealing here, I believe that those who deny that possibility are left seriously exposed by the critique in this paper. In the first place, we discovered in section 3 above that Kant’s argument is not and was never meant to be a general argument against the possibility of moral luck. It is at most an argument against the possibility of moral luck in the way that our actions turn out (3.2 above), and even then only against the possibility that the morally virtuous are exposed to moral luck in the way that their actions turn out, not that the morally vicious are (3.1 above). Those ‘Kantians’ who believe that there is a more sweeping or comprehensive problem of moral luck thus need to provide a quite new argument to show why and how this can be so. On top of this, those who join me in rejecting Kant’s doctrine of moral autonomy cannot help themselves even to the narrower objection to moral luck which Kant’s argument may be thought to yield. For without that doctrine of moral autonomy, premiss (a) cannot be established and the argument to (d) never gets going at all (4.1 above). This leaves Kant’s argument unavailable to the large proportion of moral philosophers, and the even greater proportion of non-philosophers, who hold that moral value can be found in truly spontaneous right action as well as in, or rather than in, self-consciously right action, and who therefore endorse something more like (a’). I fail to see why, for these people, any kind of moral luck is supposed to be a

problem, because I fail to see what argument against the possibility of moral luck they have in mind that is consistent with their other philosophical commitments. At any rate I have never seen such an argument.

But things are slightly different for those who do continue to adhere, in spite of what I have said, to Kant's doctrine of moral autonomy and hence to (a). Even for them Kant's argument has been thwarted as an argument against the possibility of simple duties to succeed, and hence the possibility of 'strict liability' wrongdoing. For I have shown that, once (a)'s dependence on Kant's doctrine of moral autonomy is exposed, (d) turns out not follow from (a), (b) and (c) because the word 'unconditional' is used in a stricter sense in (a) than it is in (c) (4.2 above). But the argument has not been thwarted as an argument *against the possibility of moral luck in the way that the actions of the morally virtuous turn out*. It can be rescued as an argument against this possibility by careful modification of premisses (c) and (d). The non-sequitur exposed in 4.2 above does not arise, after all, in the following adjusted argument, in which the strict test of unconditionality from premiss (a) can be carried over without objection into the later premisses thanks to the elimination of all references to duty:

(a) The only source of unconditional (a.k.a. moral) value in our actions is the good will;

(b) the good will infects not the whole of what we do but only that part of it which consists in our trying to do good;

(c') whatever there is unconditional (a.k.a. moral) reason to do is such that doing it can have unconditional (a.k.a. moral) value;

thus (d') there can be no unconditional (a.k.a. moral) reason to succeed, but only an unconditional (a.k.a. moral) reason to try (unless trying entails success).

I do not deny that, unlike the argument to (d), the argument to (d') is valid assuming, for the purposes of argument, the soundness of premiss (a). But as soon as we see (d'), we are faced with the question of why it matters. Given that (d') does not rule out the possibility of duties to succeed, what does it rule out, exactly? And with what implications?

My own view is that, with (d) out of the way, (d') has no implications worth worrying about. The widespread assumption that it does have implications worth worrying about stems from the assumption that moral reasons have some special hold on us, some special status or import, beyond that enjoyed by non-moral reasons. Thus if we fail to comply with moral reasons there are some special implications of that failure which would not obtain if the reasons in question were non-moral reasons instead. There is some sense, we are led to believe, in which compliance with morality is more important than compliance with e.g. self-interested or aesthetic or economic considerations (or whatever other kinds of considerations one classes as non-moral). Accordingly it becomes crucial to know exactly which reasons are moral reasons. In particular it becomes crucial to know whether certain types of reasons that can leave us in tricky rational predicaments (e.g. reasons to succeed in endeavours in which trying does not entail success, or reasons to do what cannot in practice be done) can be moral reasons. It becomes crucial to know, for example, whether my reasons to save the drowning man as I stand on the isolated cliff-top, or my reasons to telephone my mother on her birthday, can be moral reasons. For if they cannot be moral reasons then happily the special status or import of moral reasons does not extend to these cases, and we can thus take a somehow more relaxed attitude to the difficulties of rational compliance that they present. We need worry less about the extent to which our doing what these reasons would have us do lies out of our hands. Heaving a sigh or relief as I stand on the cliff-top incapacitated by the horror of my situation, I

may at least reassure myself with the thought: 'This is quite a predicament, to be sure, but at least it's not a *moral* predicament!'

The difficulty, however, is in understanding what special status moral reasons are supposed to have that makes this sigh of relief intelligible. Broadly, I can think of two interpretations of the idea that morality has a special status which might conceivably be relied upon here. According to the first, failure to do what a moral reason would have one do reflects personally on the agent, shows her in a bad light, while failure to do what a non-moral reason would have her do does not. Thus one is necessarily tainted by the immorality of one's action in a way in which one is not necessarily tainted by other forms of irrationality. Thus my relief on discovering that my predicament is not a moral one is the relief of thinking that my inaction will not reflect upon me personally. But if this is true, it is true only thanks to (d') itself. It is only because one can have no moral reasons to succeed, except where trying entails success, that whenever one does not do as a moral reason would have one do, this necessarily shows one in a bad light. Recall Kant's proposal, discussed in 3.1 above, that while one can be a morally deficient person without invariably performing morally deficient actions, one cannot ever perform morally deficient actions without being a morally deficient person. The soundness, for Kant, of the second limb of this proposal turned on the fact that, in his view, there can be no moral reason to succeed except where trying entails success. All of those who failed morally failed at least in part because their will, their moral character, was deficient. But if we were to abandon the assumption that there can be no moral reason to succeed except where trying entails success, we would also in the process automatically eliminate the implication that every morally deficient action reflects badly on its agent. Thus premiss (d') turns out to be in this respect philosophically self-important. Impressive though it appears, if it were abandoned nothing would be any different. It would still be the case that committing a strict liability wrong does not reflect badly on one.

My sigh of relief at the thought that at least I am not in a *moral* predicament when I am perched on the cliff's edge could well be the sigh of relief at the thought that the mere fact that I will not do my duty and save the drowning man is not for want of diligence on my part, nor for want of humanity, nor for want of any other virtue. But it matters not one jot for this purpose whether the duty or the virtue in question is regarded as a moral one. For whether it counts as moral or non-moral, no excellence or deficiency of character is reflected in the mere fact that one did not do as a reason to succeed would have one do. So far as this point is concerned, then, conclusion (d') changes nothing and can be disregarded as a red herring.

The other relevant interpretation of the idea that moral reasons have some special status is the interpretation according to which moral reasons have a special *stringency* which affects how they are to be counted in our reasoning. On one view of this stringency moral reasons defeat non-moral reasons when the two come into conflict. Morality is overriding. On another view it is merely that non-moral reasons can rationally be left out of one's reasoning whereas moral reasons must always be included.²⁶ But both of these views are rendered plausible only by the association, inherited from Kant, of morality with duty. Duties, as I have said time and again, are categorical mandatory reasons. Being categorical, they apply to us irrespective of our prevailing personal goals; in that sense we cannot rationally opt to discount them when they do not suit our plans. Being mandatory, duties exclude at least some countervailing reasons from consideration, giving them a tendency to override non-duties which increases in proportion to the absoluteness of the duty. It follows that if the moral domain is coextensive with the domain of duty, then morality has a special stringency – a special *double* stringency, in fact – which separates it from the non-moral domain. For moral

²⁶ For discussion of these and other dimensions in which morality is sometimes thought to be off-puttingly demanding, see Samuel Scheffler, *Human Morality* (Oxford 1992), 17ff.

considerations are then both mandatory and categorical. But once the connection between morality and duty has been broken, either by the recognition that there could be non-moral duties (e.g. duties to oneself, duties towards one's art, duties of love) or by the recognition that some moral reasons are not duties (e.g. reasons to perform supererogatory acts) then the special stringency of morality instantly ebbs away. For now there can be non-moral reasons which possess the relevant kind of stringency, or conversely moral reasons which do not. I for one would make this break: that there could be no moral supererogation and that there could be no non-moral duties both strike me as bizarre implications of any moral view. But that is not the point which matters just now. The point is that whether or not the connection between morality and duty holds, those who made the jump from Kant's premisses (c) and (d) to the modified premisses (c') and (d') have themselves abandoned the connection of morality with duty, because doing so is the only way to avoid Kant's non sequitur which I exposed in 4.2 above. They have at the very least been forced to accept that, according to the test of moral validity implied in premiss (a), there can after all be non-moral duties. Thus the special stringency often associated with moral reasons is extended to at least some non-moral considerations. So the significance of conclusion (d') cannot be based on that special stringency.

This leaves us, so far as I can see, without any reason to care about conclusion (d'). Why should I care that my ghastly rational predicament, as I stand perched on the cliff top watching the drowning man below, is at least not a moral predicament? With the abandonment of Kant's conclusion (d) it cannot be that I now have no duty to save the drowning man. It cannot therefore be that the reason I have to save the drowning man has no special stringency which makes my rational immobilisation so much the more vexing. For save insofar as they are duties moral reasons have no special stringency that I can think of. So wherein lies the supposed problem of moral luck isolated by (d')? Since the

problem of moral luck is no longer the problem of whether I can have a duty to succeed even where trying does not entail success, since it is not the problem created by Kant's premiss (d), what problem is it? At this point I am at a loss to see any problem at all.

It may be objected that, by denying that there is a distinct problem of moral luck captured in the argument to (d'), I am effectively reducing my own conclusion to the status of the trivial. My conclusion is that there can (or to be exact that nobody has given us any reason to suppose that there cannot) be duties to succeed irrespective of whether trying entails success. There can, in other words, be strict liability wrongdoing. But as it stands, it may be said, this conclusion has never been seriously denied, even by Kant. Why, even Nagel, who gives Kant's argument an untenably extreme interpretation as well as giving it a great deal of credence, accepts without any hesitation that there can be *legal* duties to succeed irrespective of whether trying entails success. There can be strict liability wrongdoing *in law*.²⁷ Thus there obviously can be duties to succeed irrespective of whether trying entails success, and there obviously can be strict liability wrongdoing. My conclusion goes without saying. The only real puzzle is whether that wrongdoing can be moral wrongdoing as well as legal wrongdoing. By dismissing conclusion (d') as inconsequential I am refusing to tackle that puzzle. So all that I have shown, by my long and tortuous argument in this paper, is what everyone already accepted as obvious. To get beyond the obvious, the objection goes, I have no option but to tackle (d').

This objection is mistaken. The argument of this paper tackles and resolves the only real puzzle to be found in the debate about moral luck, which is the puzzle about the existence of duties to succeed. To expose the mistake in the objection we need to make a small final detour into the general theory of normative discourse. For we need the general theory of

²⁷ Nagel, 'Moral Luck', above note 10, at 31.

normative discourse to understand one sense in which, in spite of all I have said against Kant's conception of morality, morality undoubtedly *is* unconditional. And we need to understand this sense before we can see the mistake in the objection.

Many statements invoking or presupposing reasons and values may be made with more or less detachment on the part of those who make them. Legal statements are like this. As a lawyer, I may utter statements like 'you have to do as a prudent man of business would do' or 'freedom of contract is paramount', without myself endorsing the reasons and values which I thereby invoke. I may try to make my detachment clear by prefacing the statements with caveats like 'according to law' or 'from the law's point of view'. I may do much the same thing with, for example, the reasons and values internal to a game. In the context of a discussion of football, I can say 'that was a serious foul' or 'that was a mistaken decision by the referee' while all the time believing that football is a fatuous game with idiotic rules. I utter these statements as detached statements, statements invoking reasons and values which I do not myself endorse but which are valid for the purposes of football and its discourse, a discourse in which, perhaps disingenuously, I am participating. Again I may hedge with expressions like 'in footballing terms' or 'according to the rules of football'. Likewise with many other families of reasons and values, more or less systematically organised. I may advise you on whether or not to eat certain food, given religious beliefs which you hold but which I do not share, or on how, according to the company's procedures, which I regard as ridiculous, certain kinds of complaints are to be dealt with. In this respect, however, a moral statement is different. A statement which takes, for the purposes or discussion or argument, a rational point of view which is not endorsed by the person who utters it is not a moral statement. I do not mean by this that every statement containing the word 'moral' or its cognates is necessarily a committed statement. One may utter, with more or less detachment, statements invoking the reasons and values of

'bourgeois morality', 'Kantian morality', 'moral conservatism', etc. All I am saying is that these are not moral statements. Rather, they are statements about reasons and values that some people hold to be moral. Holding reasons and values to be moral means, among other things, being committed to them, so that, when the question arises of whether one should also be committed to the reasons and values internal to the English legal system or to football or to vegetarianism or to modern architecture, what one then holds to be *moral* reasons and values are already automatically installed, so to speak, among the resources which are at one's disposal in answering the question. As rational beings we have the option to ridicule or decry some reasons and values, to ignore others, to become committed to some, and to invoke all and any of these *arguendo*. But concerning what we hold to be moral reasons and values we have, as rational beings, no such array of options. That we are rational and they are moral entails that we are committed to them.²⁸

It is true, of course, that we may make mistakes about morality, so that we become committed under the heading of 'morality' to the wrong reasons and values. But this does not affect the main point. It only adds that, since a rational being is necessarily committed to whatever she holds to be a moral reason or value, whatever she ought to hold to be a moral reason or value, because that is what it is, is a reason or value to which, necessarily, she ought to be committed. And in that sense, the appeal for and application to rational beings of moral reasons and values is, as Kant says, unconditional. For rational beings, to put

²⁸ It does not follow that there is no sense in prefatory expressions like 'morally speaking' or 'from the moral point of view'. These are open to two possible interpretations. On the first they hint at some detachment and so suggest that the considerations about to be adduced are not in fact regarded as moral considerations by the speaker but only by some others. On the second they do not suggest detachment at all but merely act as signals of convenience to indicate *which* committed considerations are being emphasised in what follows.

it another way, there is a question of why they should be law-abiding, why they should play or follow football, why they should be vegetarians, and why they should judge a development scheme in architectural rather than economic terms. But, in spite of the persistence of some moral philosophers in asking this question,²⁹ there is and can be no question of why people should be moral, of why they should care about moral reasons and values. We can call this feature the *unquestionability* of morality's appeal.

I suspect that Kant took the unquestionability of morality's appeal to imply the categorical nature of moral reasons and values. It is easy to slip into this confusion. If a reason applies to me only on condition that I have a certain prevailing personal goal (i.e. it is a hypothetical rather than categorical reason) doesn't it follow that the application to me of that reason becomes optional, so that as a rational being I may, as it were, detach myself from it? Can't I come to see its appeal, and thus the appeal of reasons which depend upon it, as questionable? So isn't a hypothetical reason necessarily non-moral by the test I just supplied? One reply might be that this mistakes our relationship with our personal goals. While we often abandon goals and acquire new ones, and at least sometimes we may opt to do so, this does not affect the rational significance that our goals have for us while they are our goals. To our goals, as to whatever we hold to be a moral reason or value, we are necessarily committed for as long as and to the extent that they remain our goals. So why cannot the hypothetical reasons which apply to us during this time qualify as moral reasons according to the criteria of the previous paragraph? Why can't hypothetical reasons be, for as long as they apply, unquestionable in their appeal? To which there is the obvious rejoinder that, to the extent that it is true that we are committed to certain reasons by virtue of our goals,

²⁹ In the case of Kai Nielsen, for example, a whole book's worth of fascinating essays entitled *Why Be Moral?* (New York 1989).

these reasons tend to strike us as categorical rather than hypothetical during the time when we are committed to them, and this allows us to explain how we might come to regard them as unquestionably appealing: we (mistakenly) regard them as unquestionably appealing just in case we (mistakenly) regard them as categorical. This debate could go on, but is doomed, alas, to inconclusiveness. The error on both sides is to try and approach the problem by considering the place of hypothetical reasons in the reasoning of the person who has the personal goals on which those reasons depend for their application. Instead one should consider the way we may think about reasons which apply to *others* depending on *their* personal goals. I may advise you, for example, that if you want to be a doctor you will have to learn to respect people more. You may ask me whether I am advising you according to the professional code of conduct of the General Medical Council, or according to the law of the land, or according to the standards set by hospitals, etc. I may respond that it is none of these. I do not know what the rules of the General Medical Council or the law or the hospital say on the matter, so I am not competent to advise according to those rules, but I certainly know what I think: I think that those who want to be doctors have to learn to show more respect for people than you show. Here I am denying that the appeal of this imperative is, for me, questionable. And yet it is plainly, as I interpret it, a hypothetical imperative. It requires that you learn to show people greater respect, but only on condition that being a doctor is a personal goal of yours. One could put this another way by saying that the mere fact that the personal goal is questionable, in the sense explained above, does not entail that the reasons which apply to those who have the goal because they have it must also be questionable. I could of course be unsympathetic to your wanting to be a doctor. I may even say that you have no reason to want to be one. Thus when I talk and think in terms of the reasons which you take yourself to have for being a doctor I may express my detachment from them by adding caveats like ‘from

your point of view' or 'as you see things' or 'if I were you'. But it does not follow that all the further reasons which this goal brings with it, such as the reason to for you to learn to be more respectful of people if you want to be a doctor, must also be reasons from which I am similarly detached. On the contrary, it may be that the very reason why I believe you to have no reason to want to be a doctor is that I believe that one only has reason to want to be a doctor if one can and will learn to be more respectful of people than you are or will ever be. It may be that the unquestionability of *that* reason – the reason for those who want to be doctors to be respectful of people – is precisely what explains, given your attitudes to people, the questionability of your reasons for wanting to be a doctor. And this shows very straightforwardly that the unquestionability of morality's appeal does not entail that moral reasons are necessarily categorical reasons.

This matters here only because it helps us to see how the pseudo-problem of 'moral luck' came into being. There is a broad sense of the word 'moral' according to which all and only reasons and values of unquestionable appeal are moral reasons and values. In this sense morality is, as Mill once put it, the whole 'Art of Life'.³⁰ Now, their appeal being questionable, legal and reasons and values are not, *qua* legal, moral reasons and values even in this broad sense (although they may of course often coincide with and reflect moral reasons, as well as *aspiring* or *claiming* to be moral). But reasons of self-interest, aesthetic reasons, economic reasons, reasons of etiquette etc. may be moral reasons in this broad sense even as they stand. Many of these have the necessary unquestionable appeal to count as moral in the broad sense. At the same time, however, there is a narrower sense of the word 'moral' according to which the realm of the moral is regarded as excluding some or all of these latter

³⁰ Mill, *A System of Logic* (8th ed., London 1872), bk. vi, ch. 12, s 6.

considerations.³¹ Kant's conception of morality, insisting that all moral considerations are categorical (and consequence-insensitive, universal, eternal etc.) is one extremely narrow conception of morality in the narrow sense. It excludes from the strictly moral realm some reasons and values which do have unquestionable appeal, but which fail the further unconditionality requirements which Kant imposes: Kant labels these 'prudential' reasons and values. His conclusion (d) has no implications for these prudential reasons, and is meant to have no such implications, since they are not covered by premiss (a). Indeed to exclude them from the picture is the whole point of premiss (a). The 'Kantian' argument against moral luck, which I schematised in the sequence (a) to (d') above, naturally shares this feature since it shares premiss (a). It is an argument against the possibility that there can be, *in the narrow sense*, moral reasons to succeed irrespective of whether trying entails success. It therefore implicates the debate about whether Kant drew the boundaries of narrow morality in the right place. Many modern moral philosophers doubt it, attributing to morality characteristics that Kant would have found problematic (e.g. agent-neutrality) and denying it characteristics that Kant would have found essential (e.g. universality). It should have been tolerably plain from my reaction to the argument to (d') what line I take on this issue. Like Bernard Williams, Charles Taylor, and Joseph Raz, I believe that the debate about which considerations are moral in the narrow sense has proved to be philosophically barren, perhaps even philosophically corrupting.³² Reasons simply have

³¹ The distinction is of course borrowed from John Mackie, *Ethics: Inventing Right and Wrong* (Harmondsworth 1977), 106–7.

³² Williams, *Ethics and the Limits of Philosophy* (London 1985), 174ff as well as (more trenchantly) 'Moral Luck: A Postscript' in Williams, *Making Sense of Humanity* (Cambridge 1995); Taylor, 'A Most Peculiar Institution', in J.E.J. Altham and Ross Harrison (eds.), *World, Mind, and Ethics* (Cambridge 1995); and Raz, 'The Moral Point of View' in J.B. Schneewind (ed.), *Reason, Ethics, and Society* (Chicago 1996).

whatever stringency they have, and values have whatever evaluative implications they have, irrespective of whether they are classified as 'moral' or 'self-interested' or 'economic' etc. How they are classified within these groups is mainly a matter of explanatory and conversational convenience. What matters for practical reasoning as well as for personal evaluation is not whether reasons and values are moral or self-interested or economic etc., but whether they are treated as having the weight and exclusionary force that they actually have when they come into conflict with other reasons. A weighty non-moral reason defeats a weak moral one just as a weighty moral reason defeats a weak non-moral one. An absolute non-moral duty excludes countervailing moral considerations as well as non-moral ones, just as an absolute moral duty does. Nothing of great philosophical significance turns on whether the reason is moral. But this comment relates, of course, to morality in the narrow sense. There is, on the other hand, great philosophical significance in the distinction between morality in the broad sense and that which falls outside it, because there is great philosophical significance in the distinction between committed and detached normative discourse. The point cannot be developed in detail here. All that really matters here is that the normative discourse of this whole paper, including the discussion of reasons to succeed and reasons to try in section 1, was all in the committed key except where explicitly signalled otherwise. Back in section 1, where I introduced the case of my predicament on the cliff-top and the case of the birthday telephone call to my mother, I was not addressing the problems thrown up by such cases in some noncommittal normative frame of reference such as the frame of English law or the frame of the Christian faith. I was addressing them as they present themselves to rational agents looking for a straight, committed answer to the question of what they have reason to do. Likewise throughout the rest of the discussion (barring certain *ad hominem* passages where I looked at certain problems purely through Kant's eyes, and hence

noncommittally, and others where I mentioned technical legal debates about the interpretation of the English law of negligence). The whole discussion, in other words, has been conducted within morality in the broad sense.

It follows that the argument of this paper supported the view that, *within morality in the broad sense*, there can be duties to succeed even where trying does not entail success, so that there can be strict liability wrongdoing. Thus I did not merely reaffirm the obvious point that there are such duties and such wrongs within some legal systems or within other specialised normative realms to which people may resort in a noncommittal way. I reached a much more significant conclusion. In the only sense of 'moral' that has real philosophical significance, viz. the broad sense, I argued that there can be moral luck in the way that our actions turn out, and that this can be so even when we are morally virtuous. What is more this moral luck can extend even to matters of duty. So even with the best will in the world, to put my conclusion in the simplest terms, we are never able to protect ourselves fully against becoming moral wrongdoers thanks to the way our actions turn out.