

Two approaches to Moral Responsibility *Garrath Williams*

Part II

Introduction

In the first part of this article (RJP Issue 6), I discussed two important understandings of moral responsibility. One approach links responsibility to the idea of free will, an idea which has, as we saw, Christian origins. On the other hand, Aristotle's account, predating Christianity by several centuries, views moral responsibility in the context of a moral community seeking to encourage and reinforce shared standards, where people make judgments of one another's character in the context of their on-going relationships with one another.

Since belief in heaven and hell is now rarely thought of, even by many who consider themselves Christian, one might suppose that our sympathies would naturally go to Aristotle's account. Yet this is not the case. The idea of moral responsibility so well captured in Kant's philosophy rests on a powerful notion of moral worth that continues to have strong intuitive appeal. Here I want to set out Kant's account in more detail, before suggesting that we can safely give up some of the intuitions that support his account of moral responsibility.

The Kantian account: freedom, intentions and control

The reason why so many people - even before they come to philosophy - feel the pull of the free will debate lies in the idea of moral worth we often associate with responsibility attributions like blame. Galen Strawson expresses the core idea as follows: 'if we have [true responsibility], then it makes sense, at least, to suppose that it might be just to punish some with eternal torment in hell, and reward others with eternal bliss in heaven' (1991: viii). Any such 'ultimate' merit or demerit is clearly a matter of strictly individual desert. If it were merely a matter of chance who went to heaven or hell - or who *would* do so, if those fates really existed - this would plainly be a matter of mere fortune. Such intense good or bad luck would make the world even more morally arbitrary than it already is. If such merit is to be fairly allocated, therefore, it needs to be seen as something that lies within individuals' own control. This line of thought, in turn, relies on what John Skorupski calls an 'ideal of pure egalitarian desert' (1999: 156). The idea is that we all equally possess such control, so that it makes sense to imagine everybody reaping an equally fair return on how well we exercise that control.

The greatest representative of this line of thought is Kant. For Kant, our moral worth - the goodness of our will - is gauged by how sincerely and persistently we have sought to do our duty. To do our duty may be much harder for some people, for instance, those who have violent passions or who were brought up with bad habits. But moral worth is not about results; it is about *the will*. We all have such a will, an ability to choose well, despite the fact that some of us face stronger counter-inclinations or more difficult circumstances. To truly judge a person's moral worth involves seeing

past all the obstacles that their will has faced. Kant argues that this makes moral worth impossible for us to judge with any assurance; only God can see beyond all those things. Kant's main concern, though, is how we judge *ourselves*. To Kant it's no problem that we're never sure about others' wills, and the obstacles or benefits they have faced. The point is that we can never be sure of our own motivations: we must keep trying to do the right thing, and to do it *because* it is the right thing.

But what is 'the right thing'? What about those who are morally ignorant? Kant denies that there is such a thing as moral ignorance, in that we are all equally well able to see what we should do - if only we try. For Kant 'even the most hardened scoundrel' would act morally, were it not for the opposing incentives of his inclinations and desires (*Groundwork*, 4:454). Kant has to claim this because otherwise he would not be able to justify condemning people who suppose they are doing the right thing, when in fact their acts are quite wicked. Adolf Eichmann, who we mentioned in the first part of this article, thought he could justify his actions (ironically, with a version of Kant's moral philosophy!). Yet no one doubts that he deserved the gravest condemnation for his crimes, even though many believe he was sincere in thinking his acts were defensible. Such examples show how implausible is the claim to equal moral knowledge. Even outside of such extreme cases, people's sensitivity to different moral considerations is highly variable, and is clearly shaped by up-bringing and environment. From an Aristotelian perspective, though, the realities of moral ignorance and moral disagreement pose no theoretical problems. In fact, they provide an important justification for praise and blame in terms of mutual accountability. Responsibility attributions are important to moral learning, by communicating how we have met or failed to meet moral standards. Because the Kantian account goes inward, to *my* scrutiny of *my* motives and intentions, it ignores this crucial *educative* aspect of responsibility attributions.

This striving to improve my intentions is tightly bound up with the idea of desert. It is a basic and very appealing intuition of Kant's ethics that *happiness should correspond to moral worth* - even though in the world we know this is invariably not the case (at least so far as we can judge people's moral worth at all). For this reason Kant claims we are rationally and morally committed to faith in a *future* world, where virtue will be rewarded. (Kant is completely silent on whether the wicked should also be punished; again, this relates to his principal concern, the striving of the self rather than the judgment of others.) Few modern Kantians follow Kant in thinking we must have faith in the compensatory schema of an after-life. But many people remain deeply attached to the idea of a distinctively *moral* sort of worth that Kant so consistently articulated.

As we have seen, such worth is measured by the sincere intentions that are within *everyone's* control. This conception corresponds quite plausibly to several features of moral responsibility. We judge the intent behind people's actions, rather than the often haphazard results of our actions. We take account of people's circumstances, and judge less harshly where these place hard or immoral pressures on people. We also, quite often, feel that allowances should be made for the effects on character of abusive or deprived upbringings. In each case, we can interpret these concessions in Kantian terms - as drawing a distinction between the person's will and the obstacles of circumstance, thus keeping our moral evaluation to what is within a person's

control. There are, however, good reasons to doubt whether this Kantian interpretation is really the best account of these intuitions.

First, it is a commonplace that we expect people to take responsibility for things they didn't intend. This is not only in those cases where we judge that someone should have formed their intentions more carefully. Certainly we judge the negligent driver who causes an accident more harshly than a driver who was careful but nevertheless caused an accident. But even in the latter case, we expect the driver to bear important responsibilities. (The problem that many of the things which attract moral merit and demerit are wholly or partly outside of individual control is known to philosophers as the problem of moral luck. Kantians *deny* that our moral worth could ever depend on luck, whether this is a matter of our upbringing or the circumstances we encounter.)

Second, when we are allocating things on the basis of desert, this is usually not to do with what the individual concerned has been able to control. The usual criterion for a job appointment is merit, measured in terms of the demands of the post: on this basis, the best applicant deserves to be offered the job. The merits concerned may be matters outside anybody's control, for instance, the looks of a fashion model, or they may be only partly related to those qualities of character that we assume people have some control over. Why, then, should control be so important in allocating '*moral desert*'? (Perhaps it will be said that it is a defining feature of moral desert that it relates only to what is within our control. Nonetheless, the question remains: why should we think this to be so?)

Not least, some of the concessions to circumstances that the Kantian story permits can be accounted for in quite different terms. We might take account of *intention*, for example, because an intended action reveals character especially clearly. But if we always look back to the will, we face grave difficulties in accounting for the moral evaluation of unintended actions. Just the fact that a person did *not* intend to do something may say something important about her character. Negligence is a case in point: the person lacks appropriate habits of attention. Inconsiderateness is another. The person does not *intend* to be selfish; rather, he *fails* to form intentions with regard to the effects his conduct has on others. In other words, the Kantian account, being focussed upon control, stumbles when a person fails to control herself. An account based on character, on the other hand, cares both about a person's intentions *and* the things she does out of mere habit or inattention.

These are reasons for caution but not definitive objections to the Kantian notion of moral worth. One might, for instance, claim that responsibility in practice represents a combination of Kantian and *consequentialist* tendencies. We want to be fair to people's moral worth, but we also need to take account of the consequences of how we allocate responsibility. Thus we usually make concessions to intentions, circumstances, up-bringing in the name of individuals' moral worth. But sometimes we have to be more concerned with overall results than individual fairness - for instance in public systems of punishment and liability. Here, there might be good reasons to punish or reward regardless of inner striving, especially as it is so difficult to know about a person's will.

The idea of worth

The notion of moral worth central to Kant's account is probably what one writer on ancient Greek ethics - AWH Adkins - had in mind when he said, 'We are all Kantians now.' (1960: 2) Kant's idea attractively reconciles two broad value judgments: (i) the moral egalitarian idea that all persons are moral equals by virtue of having freedom to choose morally; and (ii) the idea that responsibility relates to desert, so that people can nonetheless be judged very differently - some being blamed for their lives and characters, others praised. We have seen that it is not really plausible to think that people have an equal ability to choose well. But the deepest difficulties for a Kantian account turn on the idea of a moral worth, where moral responsibility is understood in terms of a person's 'ultimate' merit.

To begin with, contrast Kant with Aristotle. Aristotle makes no claims about a person's ultimate merit or demerit. People might be vicious or virtuous in various ways, and there might be rare paragons who possess a comprehensive set of virtues (yes, these are philosophers!). Naturally we would not want to associate with the vicious, and naturally we will want to condemn their vices in no uncertain terms. It might help them to learn to do better, and it may caution others against them, and it should reinforce our own and other people's sense of what character traits are desirable. But for Aristotle there is no sense that the vicious are earning a lasting form of discredit that should condemn them in the eyes of an ultimate judge. If the vicious person were to protest to Aristotle that the condemnations he faced were unfair, perhaps because his character had been formed by his vicious parents, one suspects Aristotle would be rather unmoved. Life isn't fair, he might say, and we certainly won't make it fairer by pretending some vices are less real because of their origin in early childhood, let alone because of their fixity within an individual's character. It may be unpleasant (he might continue) for you to hear this blame and condemnation - indeed, I'm glad that it is, because at least it shows that you are not so vicious that you don't care about others' opinions of you - but there are other matters at stake here, above all the standards and expectations which regulate all our lives together.

This line of thought will always be unsatisfying if we think that our praise and blame of one another reflect peculiarly deep truths about people's moral deserts - so that there are some people who are *evil* and deserve to go to hell, others who are good and deserve to go to heaven. But this way of thinking poses a series of deep problems.

In the first place, it is very difficult to know what to make of it within a secular framework, where religious faith has become a personal matter. In particular, why should someone with no religious faith continue to believe in this idea of personal moral worth? We should not forget that there are several morally *disreputable* motives that can make this idea attractive. Revenge against wrong-doers, or hatred of them, might be justified by thinking that they really deserve these reactions. And Nietzsche made a famously cynical point: if we think of ourselves as free, and also see ourselves as leading good or at least blameless lives, then we can smugly take credit for our moral superiority over others. (Nietzsche was also rightly indignant at some early Christian writers, who claimed that one of the pleasures of heaven would be looking down on hell and the tortures of the damned.)

Second, and perhaps more important, such thinking is incoherent on its own terms. Ultimate deserts are, precisely, ultimate: they are for God to judge. Perhaps a

priesthood might be thought of as having been granted some provisional right to judge, but such an idea has become increasingly unconvincing even to the devout. Again, an individual concerned with personal salvation will naturally wish to scrutinise himself to become worthy of God's esteem. But it seems over-ambitious, to say the least, to construe our praise and blame of one another as tentative or fallible versions of what God will say. (Perhaps the non-believer could say here: given that God does not exist, it is *our* task to ensure that people get their moral deserts. To this one can object: (i) As before, it is not clear why we should think that this 'ultimate' moral desert exists; (ii) Even if it does, why should rewarding and punishing it be of particular concern to us?) Although Kant's writings are littered with moral evaluations of people and their actions, his theoretical position is that we are in no position to judge the worth or deserts of others.

This leads to a third large problem in thinking in this way: it simply does not fit with what usually goes on when we hold one another accountable. In our resentment or indignation, we might want to lend our judgments as much force as possible, and so rely on any ideas that can lend our mere words greater authority ('You deserve to burn in hell!'). But when we're not over-egging the moral pudding, it is plain that praising and blaming are fairly mundane practices. Holding each other responsible assumes that we all get things wrong in one way or another, and that for various reasons some people get things badly wrong, and that we must - for just these reasons - seek or reinforce a shared set of standards and expectations. Moreover, while we do sometimes impute good or ill will to people, it is a childish morality that is preoccupied with who is 'good' and who is 'bad': our mature moral judgments are much more complex, multivalent and fine-grained than this.

Why does the idea of moral desert continue to preoccupy us?

Why, then, does the 'moral bank account' Feinberg ironically referred to, the 'true responsibility' Galen Strawson so memorably describes, continue to hold such sway over our moral sensibilities?

As we have seen, there is a *distributive* aspect to our responsibility attributions. If something has gone wrong, and several people have had a hand in the matter, we often go to great lengths to assess who bears what degree of responsibility - who should make recompense, who should apologise, or even who should be punished. At least in part, the question is certainly: who *deserves* to be held accountable? But does this necessarily point us toward the idea of lasting moral merit or demerit? The idea of a stain on our character, when we fail to acknowledge our guilt, or do something that cannot be made good, might support this way of thinking. In another way, the idea of 'recompense' suggests the image of an account, albeit one we can again make good. But for both cases a more mundane explanation is possible: that we judge something about the character of the person we are dealing with, from how they act and how they respond to others' responsibility attributions. How well does he understand the needs and interests of those around him? What sort of cooperative relations does it make sense to pursue with him? What can we trust him with? Not least, is this a person who will *take and accept responsibility*? (Note, further, that these things do not form a neat package: some people might do better on one count than they do on another - hence the many different virtues we ascribe to different people. This point

might reinforce doubts about the single, 'ultimate' evaluation of a person's will as good or bad that is so important to Kant's account.)

Naturally, to judge such questions we need to appreciate what pressures were placed upon a person when he acted, how he understood the situation he was responding to, and special factors affecting his ability to deliberate and choose. Hence Aristotle's concern with force of circumstances, factual ignorance, and intoxication, and our more modern concern with mental illness. But on an Aristotelian account, the point is that *these factors alter the extent to which actions reveal the character of the person*. That they undermine the person's 'control' is true, *but subsidiary*. We can see this by considering how certain forms of bad character constitute a lack of control over one's actions - thus the person who is weak-willed or indecisive, for example. Here weak-willed, indecisive action reveals the person, in a way that action based on ignorance of the facts does not. This also makes good sense of Aristotle's, and our, ambivalence about coerced action. Plainly someone else, the coercer, can fairly be held responsible for the coerced person's deed; and if we were concerned solely with *distributing* responsibility that might be all we needed to say. But clearly our judgments do not come to rest so quickly: we also, habitually, evaluate the coerced person's conduct. Should she have done such and such in response to such a threat? Was giving way cowardly or prudent, feeble or tough-minded?

In other words, distributing one particular quality, blame or 'blameworthiness,' is not our only or even our real concern in cases of harm and wrong. In the first place, there is the judgment of character, which is *not* an exercise in distributive fairness. (When we are judging people in terms of their fitness and abilities to interact with others, our concern is extremely partial: above all, to evaluate those people who *we* might continue to interact with.) Second, the most important thing to be distributed is not so much *blame* but the resulting *responsibilities*: who should pay compensation, apologise, or - in the case of those who have manifested the most harmful sorts of irresponsibility - be punished. Certainly this exercise is guided by the distributive ideal of fairness, but only in part. Practicality, limited knowledge, and many consequentialist considerations also play important roles. Third, as Bernard Williams has stressed, an important part of what we seek in praising and blaming is 'to recruit people into a deliberative community that shares ethical reasons' (1995a: 16). Thus when we blame we demand that someone take *more* seriously reasons to think or act one way rather than another - more simply, that he takes *our* reasons to be *his* reasons too.

Simplifying both positions, we can say that what essentially separates Aristotelians from Kantians here is whether there is something 'more' at stake in our attributions of blame. Once character has been judged, responsibilities distributed and others encouraged to take on board our reasons and evaluations, Aristotle's story is at an end. From such a perspective the question is not, 'How can people be truly responsible for their deeds?' but rather, 'Why should a notion of "true responsibility" continue to preoccupy us?'

The Kantian will feel that something has been left out. The person whose character is judged wanting, the person who should make recompense, the person who failed to appreciate a reason to act differently - mustn't they have *earned* the blame if they are truly to *deserve* it? And how could we earn condemnation except by doings for which

we are *truly* responsible? My own view, it will be clear, is that we should resist the temptations of such an 'ultimate' responsibility. We do not need such a notion to make sense of how we hold one another responsible - at least, not of how we hold people responsible when we are not being self-righteous or violently indignant. While one strand of Christianity may still tempt us to think of the righteous and of the damned, another embodies a rather deeper wisdom. Even as a person holds the wrong-doer responsible, it is possible to say, 'There but for the grace of God go I.'

Conclusion

Ideas about responsibility are usually presented in terms of a contest between two positions, compatibilism and incompatibilism. Incompatibilists accept the dilemma of free will versus determinism: responsibility depends on me controlling my actions, rather than other causal influences that operate upon me. Praise, but especially blame, make no sense if determinism is true. Compatibilists, on the other hand, want to insist that the causal well-orderedness of the universe is, precisely, compatible with our responsibility for our actions. But for most philosophers the question is *not* whether responsibility and causal well-orderedness are compatible, but *how*. In other words, to adapt Adkins's adage, 'we are all compatibilists now.'

However, the essential issue for any compatibilist position lies in the conception of responsibility it relies on - and this issue has been much less well-explored by philosophers than the metaphysics of freedom and determinism. I have contrasted two broad schools of thought that reflect large aspects of how we put responsibility into practice. When Adkins claimed that 'we are all Kantians now,' he was not referring to Kant's (*incompatibilist*) metaphysics but rather to our tendency to feel that responsibility attributions must have depth, that they reflect something about a person's 'real' deserts. Yet this position leads us to claims about control over the self, to the idea of choices that were 'really' our own and not the result of any external influence - a position that always threatens to bring us back to metaphysical freedom and to incompatibilism.

The roughly Aristotelian alternative sketched here owes much to Bernard Williams and his critique of a distinctively modern notion of 'morality,' a notion most systematically expressed in Kant's philosophy but also expressed in alternatives such as utilitarianism. Williams argues that these ideas neither make sense on their own terms, nor do they make sense of what we actually do when we do engage in attributions of responsibility. As we have seen, Aristotle's account of praise and blame is based on: (i) how far acts reveal character; (ii) the fair distribution of responsibilities to act; and (iii) the attempt to exchange reasons, share standards, and maintain relationships with those whom we judge - and who judge us in turn. This account involves no pressure to think of people as responsible for their acts in any 'deep' sense - the sort of deep sense that would make eternal punishment or eternal reward intelligible. The basic facts of compatibilism's causally well-ordered world are that we can rely on no-one to judge such deserts - except ourselves - and that we can rely on no-one to mete out such rewards or punishment - except ourselves. In this situation we must do our best to relate to each other as best we can, and do our best with one another when we do not relate to others as well as we might. Fortunately this does not depend on an untenable idea of 'true' responsibility - only that we encourage

people to take responsibility for their actions, and be prepared to do the same ourselves.

Garrath Williams
University of Lancaster

Acknowledgments

I would like to thank David Archard for comments on a version of these articles. Much of this material also appears, in rather different form, in an entry on 'Praise and blame' for the Internet Encyclopedia of Philosophy: www.utm.edu/research/iep

References and further Reading

Adkins, AWH (1960) *Merit and Responsibility*, Clarendon Press, Oxford Aristotle *Nicomachean Ethics* (the most readable translation is Roger Crisp's, Cambridge University Press, Cambridge, 2000)

Feinberg, Joel (1970) *Doing and Deserving: Essays in the Theory of Responsibility* (Princeton University Press, Princeton NJ) - a set of classic essays on responsibility for action

Fingarette, Herbert (1967) *On Responsibility* (Basic Books, New York) - another set of classic essays, including the argument that blame is intelligible only insofar as it addresses a person's pre-existing concern for others

Kant, Immanuel (1784) *Groundwork to the Metaphysics of Morals* (the best translation is Mary Gregor's, Cambridge University Press, Cambridge, 1998)

Korsgaard, Christine (1996) 'Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations' in her *Creating the Kingdom of Ends* (Cambridge University Press, Cambridge) - a sophisticated Kantian account of holding people responsible

Midgley, Mary (1991) *Can't we Make Moral Judgments?* (Bristol Press, Bristol) - an accessible argument for the importance and inescapability of moral judgment

Skorupski, John (1999) 'The Definition of Morality' in his *Ethical Explorations* (Oxford University Press, Oxford)

Smiley, Marion (1992) *Moral Responsibility and the Boundaries of Community: Power and Accountability from a Pragmatic Point of View* (University of Chicago Press, Chicago) - criticises conventional discussions of freedom and determinism, claiming that they fail to investigate the idea of responsibility

Strawson, Galen (1991) *Freedom and Belief* (Clarendon, Oxford)

Strawson, Peter (1974) '*Freedom and Resentment*' in his *Freedom and Resentment and Other Essays* (Methuen, London) - this famous essay resituates the free will debate by highlighting the importance of 'reactive attitudes' such as resentment to interpersonal relations

Williams, Bernard (1993) *Shame and Necessity* (University of California Press, Berkeley CA) - a sustained argument that the ancient Greeks had a nuanced and sophisticated account of responsibility attributions

- (1995a) 'How Free does the Will Need to be?' in his *Making Sense of Humanity and Other Philosophical Papers, 1982-1993* (Cambridge University Press, Cambridge)

- (1995b) 'Voluntary Acts and Responsible Agents,' in his *Making Sense of Humanity*