

Two Approaches to Moral Responsibility *Garrath Williams*

Part I*

Introduction

The American legal philosopher, Joel Feinberg, once observed that ‘moral responsibility... is a subject about which we are all confused.’ (1970: 37) Here I want to contrast two influential philosophical accounts of why we make responsibility attributions – for instance, by praising and blaming people, in saying that someone deserves to be punished, and so on. In these practices, we respond to other people – and ourselves – as the authors of their actions. If they act well, we feel they deserve our admiration and sometimes gratitude or loyalty. If they act badly, we will tend to resent them, and feel they ought to make up for their actions, or perhaps even be punished. And we often feel guilty, remorseful and sometimes proud of how we have acted ourselves. In short, we think of people as morally responsible.

One very influential approach to this subject is broadly *Kantian*. This view sees responsibility for actions as stemming from our ability to exercise *self-control*. On this account moral responsibility exists because a person freely chooses her actions, and tends to lead us toward the idea of free will. In addition, because praise and blame respond to the person as the chooser of her deed, they *recognise* her dignity as a rational agent, as modern followers of Kant tend to put it. A much older approach goes back to *Aristotle*. This view situates attributions of responsibility in terms of our on-going relationships with one another. This more nuanced approach stresses the importance of *mutual accountability*, *moral education*, and *assessments of character* in terms of the many vices and virtues.

I will not try to convey the exact details of these philosophers’ accounts. What I do want to show are two things. First, how their ways of looking at mutual accountability capture important parts of our everyday commonsense. One modern commentator claimed that, in our attitudes to moral responsibility, ‘we are all Kantians now’ – by ‘we’ meaning not just philosophers but all Western persons (Adkins, 1960: 2). Another central figure in this debate, Bernard Williams, agrees that Kant captured a widespread tendency of modern moral thinking, but also claims that there exist important counter-tendencies in our practices of responsibility. For Williams, ancient Greek understandings are actually more realistic and helpful than the Kantian one. I think Williams is quite correct in this, and the second point of these articles will be to suggest that so far as our ideas of moral responsibility actually make sense, they are best captured by a (roughly) Aristotelian account.

In this first part of the article, I want to sketch two things. First, I will say something about the idea of free will. The paradoxes involved in this idea often occur to people even before they come to philosophy, and these difficulties will be central to Kant’s account. But second, before turning to Kant, I would like to tackle Aristotle’s broad approach, and show that, before free will was invented by Christian philosophers,

* Part II of this article is in issue 7.

there was a quite different way of thinking about moral responsibility – one that has much to teach us.

In the second part of the article, to appear in the next issue, I ask why Kant's account continues to attract many people who would not dream of calling themselves Kantians – indeed, many who have never even heard Kant's name. Kant's theory involves a powerful idea of moral worth based on choice. This idea, though problematic because of the idea of freedom it seems to depend on, does account for many of our intuitions about moral responsibility. But it is not the only explanation of these intuitions, nor – I will argue – is it the most plausible.

The Problem of The Will

The free will debate has become an old chestnut of modern philosophy. It is an intuitively plausible way of approaching the issues – familiar to many even before they encounter philosophical texts. It is perhaps surprising, then, that this debate is actually a rather modern one.

The basic gist is this: if I am to be responsible (*really* responsible) for my conduct, then it must be within my control. However, if it is true that every event in the universe is determined by causal laws, then this must be true of the events that constitute my actions. Therefore, my conduct cannot really be within *my* control; therefore, I am not *really* responsible for my conduct. Two conclusions immediately suggest themselves. One is that it is incoherent to praise or blame me – and everyone else – for our actions, because it is so difficult to doubt the causal well-orderedness of the universe. The alternative conclusion, scarcely more appealing, is that the human will somehow sits outside this causal framework – ie, we have free will – because it is unthinkable that our moral ideas be so desperately incoherent.

Both lines of thought are *incompatibilist*; that is, they see the ideas of responsibility involved in praise and blame as incompatible with the causal well-orderedness of the universe. But while both attract some limited support among philosophers, the overwhelming consensus now lies with *compatibilism*. This is simply the thesis that responsibility and causal order are compatible. Most philosophers agree that the alleged incompatibility results from some important confusions, although there is much less consensus about what these may be. At least one area of confusion is clear, however, and forms the central issue of this article: what sort of responsibility for conduct is involved in praise and blame? Several familiar points in the free will debate are helpful for approaching this.

In the first place, it is well-known that this debate does not turn on the truth of determinism as such. Determinism is the idea that every event is determined by fixed causal laws. Yet it may well be that every event is somehow random in origin. One interpretation of quantum physics claims that causal laws are the product of statistical regularities, while these regularities stem from a near infinite number of random events. So far as the human will is concerned, this makes no difference. If my conduct is the product of chance, this makes *me* no more responsible for it than does its being generated by causal laws. The point is that if *I* am to be held responsible, then *I* must control my conduct – not causal laws, nor mere chance, nor some particular combination of the two.

Second, the free will debate bears a disquieting similarity to an older controversy. In medieval philosophy it used to be asked how God's omniscience – his knowledge of everything that has happened and will happen – could be reconciled with our being subject to his moral judgment (that is, being sent to heaven or to hell). If God knows what we will do then this seems to imply that it is already decided whether we will act well or badly. And this, in turn, suggests that it makes no sense to punish or reward us. Theologians developed various doctrines to overcome this difficulty, but few sound convincing to modern ears – perhaps because the problem itself is no longer a live one, even for most believers. However that may be, it is interesting that many modern versions of the debate seem to take at least one of the planks of Christian theology for granted: that individuals have wills that can be bad or good, usually now expressed by philosophers in terms of people's 'blameworthiness' or (less often) 'praiseworthiness.'

In this way, the modern American philosopher Joel Feinberg ironically referred to 'a moral bank account' that we carry through life, which sums up our moral credits and debits in a single figure (1970: 20). Whether or not such an 'account' makes sense, it is at least clear that the idea of 'the will' is by no means self-explanatory.

For Kant, as we shall see, it was obvious that all my choices can be summed up in a single moral evaluation, whether I have a 'good' or 'bad' will. Kant is equivocal, however, as to whether only God might make this evaluation, or whether human beings might also form reasonable opinions on the matter. But especially if we take the point of view of mutual, *human* accountability, it is far from obvious why we should believe any such single evaluation to be possible, or what role this evaluation might play in our individual or collective lives. Certainly, we usually praise and blame in terms of particular actions and particular vices and virtues – *not* a good or bad will.

Third, this way of framing the issues creates a gulf between the conduct of normal moral agents (adult human beings of sound mind) and the conduct of other creatures – animals and children. At some stage of evolution, and at some stage toward maturity, certain animals become 'free,' whereas before they had all been 'determined' in their conduct. Although it is grossly implausible that there are no relevant moral differences between the other animals, children, and human adults, it is no more plausible that the free will simply pops into existence at a certain stage of human development. Within a Christian framework this issue was less problematic: human beings, and only human beings, have souls. Thinkers have always been aware that animals can show profound care and concern for one another, but this poses philosophical difficulties given our awareness of evolutionary continuities and the fact that the Christian idea of the soul is no longer something we can take for granted. More than this, within Christianity the moral demands upon adults could be interpreted in terms of obedience to God-given laws. Yet in the modern world demands for obedience are those we make of children. Adults are expected to obey certain basic rules, but we also expect something more – a sense of responsibility that involves judgment and initiative. Despite all this, however, we tend to think there is something sufficiently distinctive about human action, so that many non-religious people find the idea of free will plausible, and almost everyone assumes that only (mature?) human beings can be responsible for their actions.

Taking the last three points together generates a further point. If the idea of the will is complex, and there is no straightforward moral dividing line between children and adults, between humans and other animals – together, these ideas suggest that a ‘will’ is not something we all straightforwardly ‘have.’ In other words: it is implausible that all adult humans have the *same capacities*, all to the *same extent*, that are involved in controlling action. One way of retaining the idea of the will might be to think of it as the *bundle of capacities* that are needed to control action in the light of moral concerns, these capacities being set only at such a level that all adult human beings of sound mind really seem to possess them. But two points need to be kept in mind about such a strategy. First, it remains the case that people will vary in how far they possess their capacities, and this variation will largely be a product of upbringing and natural qualities – that is, *not* something within an individual’s own control. Second, the sort of ultimate control over one’s moral character supposed in Kant’s or any other ‘free will’ account is unlikely to be vindicated in this way.

The Aristotelian Approach

For an analysis of the basic set of capacities needed for moral action, philosophers continue to go back to an ancient account of moral responsibility. The terms of the free will debate are new, arising with the birth of modern science, and the theological debate about free will arises only with Christian thought. But the question of responsibility for action has always been known to philosophers, and the most famous discussion of when people can be praised and blamed for their actions remains Aristotle’s. Many have noticed that Aristotle and his contemporaries saw no need to talk about responsibility in terms of free will. Aristotle asks whether acts are *voluntary*, and whether we *attribute* them to a person or to other factors. Some have ascribed this way of framing the issues to a lack of moral or scientific sophistication on the part of the ancient Greeks. However, a number of modern philosophers, most prominently Bernard Williams and Martha Nussbaum, have suggested that an Aristotelian account is actually more coherent and sophisticated than those typical of modern philosophy – and, indeed, more coherent than our modern, ‘common sense’ intuitions about moral responsibility. Aristotle *assumes* that we are responsible for our actions (so that others can reasonably praise or blame or punish us), and proceeds by pointing to various conditions that lessen or cancel this responsibility. He discusses force of events, threats and coercion, ignorance, intoxication and bad character. (He also remarks on the continuities and differences between children’s agency and that of normal adults in ways that illuminate our practices of responsibility, a point I do not consider here.) Taken together, his account shows the basic elements involved in being a person who can reasonably be praised or blamed.

The first limitation upon voluntary action that Aristotle discusses is force of circumstances. His well-known example concerns a ship caught in a storm; the sailors must throw goods overboard if the ship is not to sink (NE 1110a). In this case the action is not fully voluntary, and we would not blame the sailors for their actions. (Nor, of course, would we blame the storm: the undesirable consequence, the loss of the goods, must be chalked off as the result of natural events, for which no one is responsible.) Note that such cases are extreme examples of the force of necessity under which we always live – we are always constrained in our actions by natural

facts, although we only tend to notice this when the constraint is sudden or unexpected.

In fact, it tends to be the interference of other people which causes us the most grief – and which really causes problems for responsibility attributions. Such interference can take many forms, but its paradigmatic forms are coercion and manipulation. Regarding coercion, Aristotle's judgment is balanced. It depends on what action my coercer is demanding of me, and what threats he makes.

Some actions are so heinous that we should be blamed for doing them, whatever we are threatened with (and whatever blame also attaches to our coercer) – thus Aristotle dismisses the idea that a man might be 'compelled' to kill his mother (NE 1110a). This makes it clear that a central issue at stake in attributions of responsibility is the *expectations* that people have of one another. There are some forms of coercion we do not usually expect people to resist, but there are also some sorts of action that we think people should never undertake, regardless of such factors. In such cases praise and blame are clearly working *to clarify and reinforce these expectations* – in other words, they provide for a form of *moral education*.

Aristotle does not comment on manipulation, where other people support us in a false view of our circumstances. But he does discuss ignorance of these circumstances, and how it undermines our responsibility. If we are ignorant of who someone is, for example – as was Oedipus, who did not know that the old man obstructing him was actually his father – we may commit heinous acts we would otherwise abhor – thus Oedipus committed patricide, killing his own father. For Aristotle, such actions are not to be blamed (at least, when the ignorance is not itself culpable and the killing was otherwise justified).

What decides good or bad character is how a person reacts when he finds out the truth – if we fail to regret our deeds, then we can certainly be blamed, even if the original choice was justifiable. Among other things, Aristotle makes it clear that our praise and blame is often not about an individual act, but about the *character* of the person who acted.

Importantly, it is not every form of ignorance that excuses. Moral knowledge is very different from factual knowledge. What if a man did not know murder was wrong? Would this make his murders morally innocent? Aristotle says not: there are certain things we can and do expect people to know – above all, basic moral truths such as the wrongness of murder. But this knowledge is not as straightforward as it might appear: it must include a fairly good capacity to judge which sorts of killing count as murder. Nazi bureaucrat Adolf Eichmann organised the killing of thousands, without a sense of its wrongness. Aristotle is clear: such moral ignorance, an inability or failure to judge, excuses no adult. Eichmann should be held responsible for murder. But why should moral ignorance not excuse, when factual ignorance does? We must recognise that moral knowledge is actually rather different from factual knowledge. *If a person is morally ignorant it is his whole character, his lasting ability to judge and act well that is impaired* – and presumably very difficult to set right. Isolated errors in factual knowledge, on the other hand, can be easily corrected. So long as we are subsequently able to recognise and regret what we have done, *factual mistakes involve no lasting corruption of character*.

Still, if a person *is* morally ignorant it follows that they are unable to choose well; and Aristotle concedes that many people of settled bad character – be they morally ignorant or otherwise – can no longer choose to act well. Does this mean that blame is incoherent or misplaced? He claims not. Even if the vicious person cannot now choose to act otherwise, there was a time when her vices were not fixed, when she *could have chosen* not to be vicious. Therefore, Aristotle says, she can be blamed. This is neat but rather unconvincing. Aristotle is famous for emphasising the importance of good upbringing and habituation, and presumably many vices are formed in childhood, before people have formed capacities for deliberating reasonably. Indeed, many vices undercut the capacity for rational deliberation. So it is a clear implication of Aristotle's own account that the badly brought up person may never be in a position to choose *not* to be vicious. Note, further, that this unconvincing move represents Aristotle at his most *Kantian*: blame is justified by reference to control, to a 'could have done otherwise' – even when his own account of character formation suggests that such control may well never have existed.

It is also interesting that many vices take the form of moral ignorance – of not knowing that certain things are wrong or failing to recognise that certain actions represent some sort of wrong-doing. (This is often a failure only with regard to one's own actions: Bishop Butler once observed how common it is for people to condemn others for the vices they themselves are most notorious for!) The difficulty is that the vicious person cannot, or will not, see her own vices as such – in which case she is in no position to 'take control' and will see no reason to act differently in the future. But this does not mean that we have no reason to blame her, most obviously because we might hope that blame will help educate her, morally speaking.

What are we to say, though, where a person seems incorrigible – quite settled in some particular vice, either because she cannot understand the criticism or because she is unable to alter her character or habits? (In real life it's often somewhere in between: 'Yes, I know I shouldn't behave like that, but I can't help it, and it's really not as bad as all that.') Such cases are very common, and – unless we suppose that they are not morally deplorable – seem to undermine the modern, Kantian assumption that blame must relate only to conduct under our control. Clearly, if we think a character trait is really beyond alteration, by us or by the person concerned, our blaming won't involve an attempt to reason with the person we condemn. But our condemnation might have another rationale, for example, to clarify what sort of standards we expect of *others*. And it is clear that praise often has this rationale, too: a virtuous person might be quite *unable* to do certain things – commit cruelty, for example.

In sum, Aristotle's account is not entirely self-consistent. Generally his focus is upon the qualities of character revealed by acts, in terms of our overall moral expectations, and it is these that responsibility attributions attend to. However, he sometimes suggests that bad qualities are to be blamed because they are, or were, subject to choice, even though this quasi-Kantian claim is not really supportable. Despite this, philosophers have returned to his account again and again to illuminate the main ingredients of responsible agency:

- The capacity to respond to others' censure and encouragement, whether expressed emotionally (eg, as resentment), institutionally (eg, as punishment), or in the various forms of praise and blame.
- The capacity to exercise deliberate, sustained control of one's conduct. One reason why young children are not responsible agents is their inability to sustain control over time, partly owing to a lack of emotional self-understanding. (That we do praise and blame children, however, emphasises the *educative* and *encouraging* role that praise and blame can play, both in developing such control and in inculcating shared moral standards.)
- A reasonable grasp of how actions impinge on others and how they are socially understood – that is, of our mutual moral expectations.

Taken together, these capacities allow us to participate in forms of mutual accountability, whereby we inculcate and to some extent enforce shared standards of action. This list may not be comprehensive, but it serves to illustrate the underlying point of an Aristotelian account: moral responsibility seems to rest on these sort of fairly basic capacities, which do not seem to demand any strong metaphysical elaboration. Indeed, if we approach the matter this way, the puzzle seems to be inverted. Not, 'how might free will and determinism be reconciled?'; rather, 'why should we feel there is a metaphysical issue at all?' It is to this question, why so many people feel that metaphysical issues are involved here, that I will turn in the second part of this article.

Garrath Williams
University of Lancaster

Acknowledgements

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, A.W.H. (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 *Can't we make moral judgments?*

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.

Williams, Bernard (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).

Williams, Bernard (1995b) 'Voluntary acts and responsible agents,' in his *Making sense of humanity*.