Responsibility, Rationality and Judgment
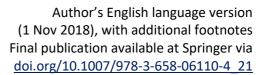
Garrath Williams (Lancaster University, UK) – g.d.williams@lancaster.ac.uk

Original publication and acknowledgements

Abstract

This chapter examines the philosophical grounds for linking responsibility with capacities to reason and to judge in the light of moral considerations. It discusses five different accounts that connect responsibility and rationality, the work of: Susan Wolf, R Jay Wallace, the jointly authored work of John Martin Fischer and Mark Ravizza, Angela M Smith, and Pamela Hieronymi. Through these authors' contributions, the chapter argues that the notion of rational ability is central to understanding and justifying practices of responsibility. Although there has been clear progress in debates about this connection, however, understanding the notion of rational or moral ability still poses profound challenges. One reason for this is suggested: such abilities may have constitutive connections with practices of holding responsible and of taking responsibility – connections that have yet to be fully explored in the literature.
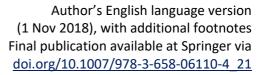
1   Introduction

Many approaches to responsibility give pride of place to human capacities for rationality. Usually this is in terms of the capacities that render an individual person responsible for her actions. Although it will not be the main focus of this chapter, there is also a wider question about the rationality – or otherwise – of the practices by which we treat one another as responsible agents.

The capacity for judgment has played a less prominent role in debates, but it is also of central importance. Judgment is involved when the responsible person grasps her situation, deliberates about what she might do, and actually acts. As some authors also argue, judgments are manifested in our spontaneous responses, our patterns of attention, and what we fail to notice. And of course, judgment is involved when people look back on someone's actions and their outcomes, and decide whether or in what ways she is responsible for them, and whether or in what ways they will respond in their turn.

For the purposes of this entry, it is useful to pair rationality and judgment, since not every philosopher understands normative considerations in terms of rationality. I will use 'normative' as a catch-all term, to capture the demands of prudence, morality, and social roles. While the demands of prudence are matters of rationality on every account, it is not clear that there is any "rational" answer as to how a person should weigh the short- against the long-term. Many philosophers doubt that morality, and perhaps still more, social and institutional expectations, can be interpreted as based on reason or our capacities for rationality. However, we can take two points as agreed: It is possible and sometimes necessary for responsible human beings to engage in reasoning about normative requirements, be they prudential, moral or matters of social expectation. And however little weight a moral theory places upon rationality, it cannot deny the central role that judgment plays in practical life. In short, we can use "rationality and judgment"
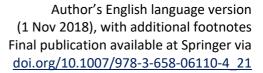
to point toward the role of moral and other normative considerations, as responsible persons act and interact.

This chapter will proceed as follows. It begins by noting the obvious reasons for connecting rationality and responsibility, and by reminding the reader of Harry Frankfurt's famous account of free agency – an account that fails precisely because it omits rationality and moral judgment. The main body of the chapter then discusses five approaches to responsibility that have highlighted the rational capacities of the responsible person, in roughly chronological order. Separate sections consider the views of Susan Wolf's *Freedom within Reason* (1990), R Jay Wallace's *Responsibility and the Moral Sentiments* (1994), and John Martin Fischer and Mark Ravizza's co-authored book *Responsibility and Control* (1998). Finally, the last two sections consider the work of two authors whose views overlap considerably, Angela M Smith and Pamela Hieronymi, both of whom have written a series of important papers about responsible agency over the past decade or so. As we will see, there are compelling rationales for connecting responsibility with abilities to reason and judge, but the requisite notions of ability raise profound issues about the nature of moral agency and moral community.

2 Background: overarching motivations for connecting responsibility with rationality (or, more broadly, with normative considerations)

The connection between rationality and responsibility is clearly fundamental. Practices of holding responsible are distinctive of relationships between adult human beings. These practices depend on our unique ability to give account of our actions, and to ask others to do this, and (of course) to reject some actions and accounts thereof as unjustifiable. The connection is reflected in the words we use: 'respond' in the case of 'responsibility,' 'account' in the case of 'accountability,' 'answer' in the case of 'answerability.'
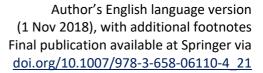
This is not to deny that such possibilities have some precursors in the relations of other animals. Other intelligent animals are liable to welcome and even reciprocate actions that benefit them, or to take anger and even retaliate in the face of actions that harm them. But since the other animals lack capacities to articulate what they have done and why, or to explain why they object to another's actions, or to articulate normative demands, we can only speak of responses, not responsibility; and we can neither speak of answers nor answerability, neither accounts nor accountability.

The basic connection between reasoning and responsibility also highlights the fundamental importance of normative standards to responsibility. It is true that we sometimes say that a person was responsible for a deed or an outcome without yet passing any normative judgment, moral or otherwise. In this case, we mean only that *she did it* or *she brought it about*: the deed or its outcome are reflections of her agency, rather than some compelling necessity or undermining derangement or bizarre coincidence. (This does still involve a normative judgment of sorts. We judge, or at least assume, that the person possesses whatever capacities are necessary for responsible agency. Still, we need not judge doer or deed or outcome as good or bad.) For the most part, however, when we say that a person is responsible for a deed or outcome, we invoke standards that govern actions: to judge someone responsible usually involves an evaluation of both deed and doer – negative in the case of blame and punishment, positive in the case of praise and reward. As indicated, whether these standards of evaluation ultimately derive from Reason is a question that we can leave aside for the purposes of this chapter. The key point is that practices of responsibility usually presuppose moral and other standards; these are standards that people can and do reason about; it is always a matter of judgment as to how they apply in any given situation; and we can also reason about these judgments.

Although fundamental, the connection between rationality and responsibility has often been overshadowed by debates about responsibility and determinism. These tend to
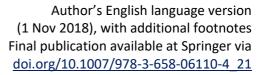
speak, rather loosely, of the freedom or ability "to do otherwise." One thing all parties should agree is that this capacity needs to be specified. Saying that a person is not 'determined' to act or think or judge in one way rather than another will not, as both Hume and Kant long ago pointed out, help us to understand responsible agency. Perhaps the universe is indeterministic, or there are particular physical configurations where outcomes are undetermined. Whatever it is, however, responsible agency is not simply random: the person must be understood as somehow self-determining. In other words, however we relate freedom to responsibility, there is a no less profound worry about how we are to understand persons as the authors of their deeds and the reasoning and judging that underlie those deeds.

Before proceeding to recent debates, it is also worth noting another argument that stands in the background for most of these authors. Harry Frankfurt (1971) once advanced a view of human will based on our capacity to identify with particular desires – the idea being that we can form "second order desires" whereby we "want to want" certain things or, of course, "want not to want" other things. The human will is essentially the ability *to want to act* on a want. For example, the willing parent "wants to want" to perform most of the duties bound up with parenthood, while the reluctant addict "wants not to want" to act on her addiction. Frankfurt says very little about responsibility, but the basic idea is that freedom of the will arises as the harmony between first- and second-order desires, and not as a matter of causal non-determination.

While Frankfurt's account may appeal as a non-metaphysical account of freedom, it faces decisive objections. One problem is that it tells us nothing about responsibility: I might "want not to want" to spite my ex-lover, for example. But that 'second-order desire' does not remove my responsibility if I nonetheless act jealously. Indeed, my incompatible desires may well lead me to self-deception rather than self-restraint – as when I rationalise my spiteful interference as legitimate concern. (Frankfurt obscures this point by focussing on examples of addictive or compulsive desires, which do seem to
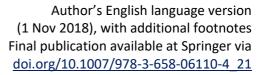
5

undermine responsibility.) More important in our present context, however, is a second problem. As Gary Watson argues in his paper, 'Free Agency' (1975), Frankfurt's account fails because it does not admit the place of reasoning and values. The question facing a responsible person is not a factual one (Does she want to have particular desires?), but a normative one (Should she act on those desires or repudiate them?). As Watson puts it, people 'do not (or need not usually) ask themselves which of their desires they want to be effective in action; they ask themselves which course of action is most worth pursuing. The initial practical question is about courses of action and not about themselves [= the desires that they have]' (1975: 219). In other words, Frankfurt's theory omits the *rational activity* that responsible persons are capable of. It fails as an account of both freedom and responsibility.

One concern of the five approaches considered below is to give a place to rational agency, without interpreting it in a way that is essentially incompatible with a deterministic picture of the world. This is not primarily because the authors concerned endorse determinism, but rather because of the point already argued: mere non-determination does not help us to understand the distinctiveness of responsible agency. While all the approaches considered below make contributions to this problem, I will suggest that there has been important progress in these debates, and that we come closer to satisfying understanding of the issues in recent work by Angela Smith and Pamela Hieronymi.

3   Susan Wolf, *Freedom within Reason* (1990)

Wolf presents her account as a response to two alternative views of moral responsibility. The 'real self' view is recognisably Frankfurt's, although Wolf also considers how it might plausibly be modified in the light of Watson's point that the actor's values are no less important than her desires. On this view, 'a person is free and responsible for his behavior when, and only when, his behavior can be governed by his will and his will can be governed by his desires… The kind of freedom necessary for responsibility [is]… the
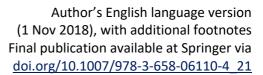
freedom to do what one's core, deep, or real self wants, which may be different from what one's strongest desires would urge upon one… [Thus] one's status as a free and responsible being lies not in whether but in how one's actions are determined' (quoting Wolf's later summary of her book – 2005: 265). By contrast, the 'autonomy' view holds that agents 'are responsible just insofar as they are autonomous. If their actions are governed by things external to their selves or if their selves are themselves governed by things external to them, then they are not responsible for the actions that ensue' (2005: 261). Clearly, such a view often motivates, if it does not already amount to, an incompatibilist view of responsibility.

With regard to the autonomy view, Wolf's objection is partly familiar. The autonomy view rejects the idea that a person's actions, along with everything else about the person herself, are determined, on the basis that this cannot allow for individual responsibility. But if autonomy is simply the non-determination of a person's thoughts and actions, then it is hard to see any sense in which a person stands behind the actions (1990: 14). Wolf sharpens this point by noting that, taken to its logical conclusion, this view implies that autonomy should be equated with the ability to choose on no basis at all – that is, regardless of what reason or desire might suggest (1990: 53f). But what, asks Wolf, could be the value or importance of such a nihilistic capacity?

Wolf's objections to the real self view are complex. Clearly important to her overall position is an argument she made in an earlier paper (1987; see also 1990: 37f, 44, 75ff; 2005: 266ff). Wolf suggests that a person who is socialized into repellent values, in such a way that he cannot appreciate reasons to act and think differently, is not in fact responsible for the resulting thought and conduct – although these patently represent his 'real self.' We should not blame such a person because he, deep down, is simply not 'sane' – that is, he lacks 'the ability to know the difference between right and wrong' (1987: 56). Her second main argument against real self views is that they cannot account for the 'depth' of judgments of responsibility. To see people simply as expressing their
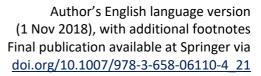
'real selves' in action (that is, when they act voluntarily rather than being subject to compulsion) would be to interpret persons simply as 'bad act-makers' or 'good act-makers' or somewhere in between (1990: 39). But judgments of responsibility do not involve merely grading people (to use a term from Smart 1961): they are 'more focused, noninstrumental, and seemingly more serious' (1990: 41). Real self views, Wolf suggests, could only address this problem if they could explain how people might be responsible for their selves. Yet this is ruled out by our sense that people are formed by their upbringing and societies, not to mention the difficulties posed by the autonomy view.

Wolf proposes that both views, because they hinge 'on a property that can be stated in a way that is not implicitly value-laden [freedom from causal determination, the ability to act in ways that one wants or values subjectively], miss the crucial feature that distinguishes responsible beings from others' (1990: 77). Instead, we should focus on people's capacity for reason – our ability to track 'the True and the Good' or 'to do the right thing for the right reasons.'[1] In terms of autonomy views, the point is not that responsible action should be undetermined by the causal influences that formed a person. Rather, actions should be determined by a person's ability to think and act in accord with reason. These two sorts of 'determination' do not compete, Wolf argues, because they involve quite different kinds of explanation (1990: 72). In terms of real self views, the point again is that responsible action cannot be understood as the mere result of whatever character traits or capacities a person happens to have. Unless those traits and capacities are connected with an ability to appreciate and act on moral considerations, responsible agency is lacking. Only persons who have some capacity to understand right from wrong are candidates for moral responsibility.

---

[1]   Despite this Platonic and Kantian language, Wolf's concern is with people's ability to think and act in terms of rationally defensible value commitments. She emphasises that her view of reason is pluralistic, presupposing only a weak objectivity about its demands (1990: 118ff).
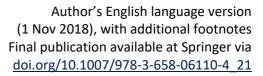
As Wolf emphasises (1990: 79ff; see also 1980), this creates an asymmetry in attributions of responsibility. There may be persons steadfastly committed to certain sorts of evil who, because they are unable to appreciate reasons to think and act rightly, are non-responsible. But persons who are steadfastly committed to 'the True and the Good,' and who simply could not (for instance) betray or steal, are still genuinely praiseworthy. The relevant test is not 'ability to do otherwise' but 'ability to act rightly.' This supposed asymmetry is controversial on both sides. We will return to the case of wrongdoers who are unable, or less able, to act well in a moment. But what should we say of the person who consistently acts rightly and even claims that he couldn't have done otherwise? If so, would he therefore not deserve praise? I will not try to settle this question, except to note (with Wolf) that it would be odd to think that an ability to do the right thing would be less morally salient if it were totally reliable. Of those rare persons who we are sure that we could count on, come what may, it is hard to feel that anything but praise and gratitude are warranted.[2]

In any event, it should be clear that Wolf identifies a fundamental point when she stresses our abilities to tell right from wrong and to reason about the difference.[3] However, it is not clear that attributions of *partial* inability should guide practices of holding responsible in the way Wolf supposes. It is not controversial that a few human beings are so deeply oblivious to rational considerations that it makes no sense to treat them as responsible – just as non-rational entities are not responsible either. But there are

---

[2]   See further Williams 1993, as well as the discussion in section 7 (below), on whether an inability to meet moral demands should exculpate a person.

[3]   Given the focus of this chapter, I will not take up Wolf's influential suggestion that judgments of responsibility have a special 'depth.' The point also arises with regard to Wallace's emphasis on the reactive emotions, briefly discussed at the start of the next section, and is sometimes presented as a worry about 'attributionist' accounts like Angela Smith's: see section 6 (below) and Smith 2008.
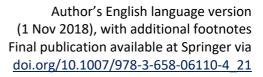
reasons to doubt Wolf's governing intuition, that those who are less able to act well are thereby 'less responsible' for their actions.

In the first place, and as Wolf concedes (1990: 87), the attribution of moral (or rational) ability is uncertain and problematic. To be sure, there is a sense in which the well-social-ised racist is not 'able' to spontaneously offer or endorse anti-racist arguments. But one might think that in most cases, there will be some ways in which such reasoning 'could' gain purchase with him – be it arguments from fairness, or compassion in the face of suffering, or what-have-you. Since the components of egalitarian argument are basic enough, ability and inability are not simple to judge.

Assuming that we can at least make *relative* judgments about moral abilities, however, note that Wolf's supposedly intuitive claim has counter-intuitive implications.[4] Morally 'less able' persons will act more badly, more often. In other words, the persons who we most need to hold responsible are the persons who we should (on Wolf's framing of the issues) regard as less responsible.[5] To recur to the previous example: if someone is utterly impervious to anti-racist arguments – immune to egalitarian argument, blind to the humanity of some persons – it is not clear that this has any excusatory force. It might

---

[4] Although rehearsing intuitions has become a popular sport in contemporary philosophy (as we also see with Fischer and Ravizza's work, below), as a method it is obviously defeated when intuitions clash. In addition – to cite just one of the epistemological issues that intu-itions pose – there is the problem of whether the questions or examples are used to 'pump' them (Dennett 1980: 429) are apt and well-framed. In the present case, we might doubt whether Wolf's question – (how far) is a person 'really' responsible? – is sufficiently clear.

[5] We might also bear in mind that many vices make it less likely that practices of holding responsible will have constructive effects. Shamelessness, hard-heartedness, self-decep-tion, aggression and so on – all make it less likely that a person will heed blame, and indeed that others will take the risks involved in voicing their blame. In other words – and here I follow a point that Barbara Herman has made in another context (2007: 266) – what looks like a theoretical difficulty (are those persons responsible?) might be more aptly framed as a practical one (how are we to interact with those persons?).
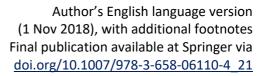
just as easily represent a peculiarly deep-seated vice, and all the more blameable for that.

A related problem arises when we consider the fact that many persons act decently just so long as they live among decent social institutions that reward (as it is sometimes said) 'pro-social' behaviour and impose sanctions on its opposite. When those institutions collapse or are usurped by tyrannical political forces, however, those same persons may become willing perpetrators of evil – betraying or killing their former neighbours, for example. History teaches us that many people will do this. Perhaps we should maintain that those persons "could have" resisted evil. (If one holds that people can only be culpable if they had the ability to do better, then such ability attributions may seem compelling.) Alternatively, we might argue that the abilities involved are not so much the property of the individual, but rather the property of the individual-in-a-given-range-of-social-contexts. This point is well-developed by Vargas 2013 (especially ch. 6).

I will not try to decide between these possibilities here – the key point, again, is just that the concept of ability is not clear-cut and poses problems of empirical attribution. But I do want to note one important implication of the latter view. It suggests that (many?) people's ability to track moral considerations is, at least in part, constituted by the fact that other people are ready *to hold them responsible* – to react with reproach and resistance if (for example) they attempt to harm their neighbours; to interact with them approvingly if they show proper consideration for others. If moral ability is partly formed and maintained by practices of holding responsible, then clearly it cannot serve as an independent basis for justifying those practices.[6] Indeed, we might argue that assertions that all sane adults have the capacity to refuse complicity in evil (at least excepting certain extremes of terror) are not simply factual claims, but also have a constitutive

---

[6]   I should add that this may not be a fair way of construing Wolf's account; indeed, much that she says about the social conditions needed for our rational capacities to develop and flourish would be consistent with the points made here.
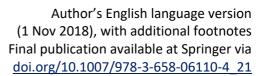
dimension. That is, such claims help to create a sense of ourselves and what we 'can' do; in particular, they help us to anticipate others' readiness to hold us responsible if we fail to resist evil. If claims about moral abilities play a role in constituting them, an unexpected meaning would lurk in that old and slippery maxim, 'Ought implies can.'

4 R Jay Wallace, *Responsibility and the Moral Sentiments* (1994)

A second approach that emphasises rational capacity owes to R Jay Wallace. His book combines two distinct strands of thought. First, it draws on Peter Strawson's famous paper, 'Freedom and Resentment' (1962), which emphasised the significance of the reactive emotions (such as resentment, indignation and gratitude) in normal human relations. Strawson argued that practices of holding responsible are not tools that we use to manipulate one another, in the hope of promoting behaviour that we approve of (thus utilitarian accounts of responsibility – e.g. Smart 1961). Rather, they reflect the profound personal importance of being treated decently by our fellow human beings. Practices of holding responsible have an emotional core that reflects our attachment to moral standards and to our own place in the 'moral community.' Wallace marries this emphasis on moral emotions and human relationships to a second, more Kantian strand. He underscores the importance of fairness to practices of holding responsible, and its link to rational capacities: 'It is fair to hold people morally responsible if they possess the rational power to grasp and apply moral reasons, and to control their behaviour in the light of those reasons' (1994: 1). Or alternatively: 'The "can" that matters in moral responsibility is not the "can" of alternative possibilities... but rather the "can" of general rational power' (7f). For obvious reasons, I will focus on this second strand.

Wallace emphasises questions of fairness because it is these, above all else, that have motivated incompatibilist approaches to responsibility. The basic incompatibilist thought is familiar: if persons are causally determined to act and think in particular ways,
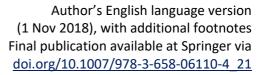
then it is not fair to respond to them with modes of holding responsible that cause them harm or pain; it is not fair because the absence of alternatives means they lack an opportunity to avoid such adverse responses. Wallace aims to do justice to this line of thought while denying that it poses a global threat to practices of holding responsible. First, he suggests that the opportunity to avoid blame or resentment should not be construed metaphysically (and mysteriously) as an ability to 'do otherwise,' but rather in the more familiar terms of rational capacity. The 'power to grasp and apply moral reasons' is a normal everyday ability. Of course, most adults sometimes fail to judge or act as morality requires; but so long as they regularly act well, we may speak of their having this ability. Second, he gives an account of excuses and exemptions – familiar cases where we do not hold people responsible, even though their activity is wrong or harmful in some respect.

Excuses and exemptions are important to Wallace partly because the standard incompatibilist argument can be interpreted as follows: *we are all excused* for our misdeeds, because given our history, make-up and circumstances, we could never have done anything else except what we did; or alternatively, *we are all exempt* from moral requirements, because we do not, in fact, have the ability to act in moral terms – granted the truth of determinism, it is merely a fortunate accident (so to speak) when a person does act well. Wallace effectively points out that this is not how excuses and exemptions work.

When persons are *exempted* from moral requirements, this is not because we suppose their actions were 'determined,' but because their deeds were somehow not subject to rational control. It would, Wallace contends, be unfair, or more specifically *unreasonable*, to hold such persons responsible. The kleptomaniac, for example, is subject to a compulsion, whereas the person who breaks a promise for personal gain is not. More seriously, the psychopath lacks capacities of rational control, such that practices of holding responsible no longer make sense. Moreover, the price of the psychopath's
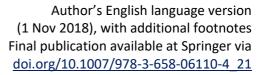
incapacity is not a welcome freedom from the costs of being held responsible, but an exclusion from the moral community. Such a person must be dealt with, as Strawson had it, 'objectively,' that is, as a dangerous entity who (or 'that'!) needs to be managed or, perhaps, treated. More modestly, the various compulsive disorders need to be managed or treated or worked around: while we do not blame a person for them, we do expect her to recognise such problems and, perhaps, to accept help to address them or minimise their impact.

Excuses operate in a more immediate and local fashion. We excuse a person or action when we suppose that, however harmful the deed and however wrongful it appears, the person did not intend to do something wrong. I will be excused for dropping a valuable vase, for example, if there were no other way to prevent someone cracking his skull against the fireplace as he faints. If a colleague who is weighed down with anxiety fails to greet us, we understand that this is nothing personal and readily excuse her. Again, the issue is not absence of alternative possibilities, in some metaphysical sense, but rather the absence of an intention to do wrong. Wallace sees excuses as being justified by the demands of fairness – specifically, that the person does not deserve blame or sanctions. Either she did not, in fact, act wrongly, or it would not be fair to insist on certain obligations given the other demands or pressures to which she was subject.

In passing, it is worth noting a problem with Wallace's argument here. It is not clear that fairness represents the 'master norm' governing practices of holding responsible (Benson 1996). One might, for example, argue that respect for persons or concern for moral community also play a role in justifying practices of holding people responsible – and indeed, some passages in Wallace's book support such thoughts.[7] One might also suggest (as Pamela Hieronymi does: see section 7 on her views, below) that it is a mistake to see blame as a sort of cost or sanction, whose distribution raises questions of fairness
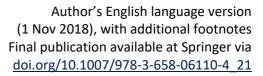
---

[7]   See footnote 9.

in the first place. Nonetheless, and as mentioned, since concerns about fairness often motivate incompatibilist views, Wallace's emphasis is understandable, even if it leaves open wider questions about the nature and justification of practices of responsibility attribution.

Behind the question of fairness lies a fundamental aspect of Wallace's approach. Wallace begins with the question of whether it is fair for people to treat one another in particular ways. That is to say, he does not assume – as many authors do – that there is some metaphysical fact about sane adult human beings that justifies practices of moral responsibility.

Evidently, facts about people are relevant to Wallace's account: in the case of exemptions, whether a person has the capacity to grasp and apply moral reasons; in the case of excuses, whether the outward deeds of a morally competent person actually reflect wrongful intentions or attitudes. But his line of argument starts with the question of how we should treat one another, and supposes that practices of holding responsible (or excusing or exempting) must be justified in those terms. Although Wallace does not put it in quite this way, this reveals a serious oddity lurking in moral arguments for incompatibilism: On such arguments, it seems that we have a moral obligation not to treat people as responsible for their actions, because every instance of immoral action is just one more instance of a person's incapacity to do anything except what she in fact does. And how, then, is the incompatibilist meant to respond to those persons who persist in holding other persons responsible? Presumably his response must be critical – and then the task is to say how that criticism is distinct from blame. (In this regard, it is striking how little has been said about the nature of blame in debates about responsibility, a neglect that has begun only recently to be addressed: e.g. Sher 2006; Coates & Tognazzini 2012. We will return to this point below, in discussing Hieronymi's work.)

This is not, of course, to say that Wallace's normative approach is definitely correct, or represents the only way to relate practices of holding responsible to wider moral norms.
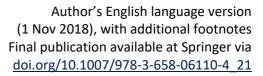
15

(In addition to the crude utilitarian approach mentioned in passing above, note also the contractualist views taken by Scanlon 1998 and Lenman 2006, as well as the Kantian view of Ripstein 2004. In different ways, all suggest that practices of holding responsible must belong to fair terms of interaction.) But critics might, perhaps, push Wallace further on the notion of capacity that plays such a central role in his theory.

Incompatibilist authors, for example, have objected that justifying practices of responsibility attribution on the basis of our rational powers will not do, if a person did not have the 'ability' to exercise those powers on a particular occasion. (This contrast between general power and specific ability is made several times in Wallace's book: e.g. 182f, 192, 199ff.) The incompatibilist worry is easy to see: if a person did not exercise her general power, then presumably she was causally determined not to do so; so reference to her general powers does nothing to vindicate a specific ability to have done otherwise.[8] Wallace attempts to pre-empt this doubt (1994: 182ff) by repeating his general argument about exemptions from responsibility: that they operate not by reference to causal determination, but by highlighting impairments to the power to govern oneself in the light of reasons. In addition, Wallace points out that the modal notions surrounding practical difficulty (such as lifting a heavy weight or speaking a second language) are quite distinct from those involved in judging that something was determined to happen and could not have been otherwise (in reply to Kane: 2002: 717ff).

One may feel some dissatisfaction with these replies. The first may look as if Wallace is simply digging in his heels, instead of giving a fuller account of how rational powers and practices of responsibility should be set alongside one another. Although he makes occasional comments that hint at a deeper connection,[9] the emphasis he gives to excuses

---

[8]   E.g. Cullity 1997, Kane 2002. For a recent attempt to address this doubt see McGeer & Pettit 2015.

[9]   Thus one of my favourite lines of Wallace's book, on its penultimate page: 'The conditions of responsibility that I have identified in this book describe an ideal… of a community of
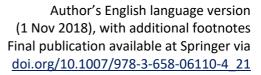
and exemptions leads him, I would suggest, to underplay people's role (when not ex-cused or exempt) as *participants* in upholding moral community. And while the second reply may address the problem as his critic frames it, it does not answer the deeper problem of just what sort of modal categories are required to understand moral agency. While we sometimes blame people for not exercising enough determination or will-power – thus the modal categories of practical difficulty – this is not necessarily the typ-ical case. Bad people do not merely have to try harder to be (or do) good, and being good is hardly analogous to straining one's muscles. If we saw matters in this way, we should no doubt conclude that some people's 'moral muscles' are simply weaker, no matter how much effort or 'training' they put in, and this would again raise worries about the fairness of responsibility attribution. (Indeed, this seems to be the way that Wallace frames matters in the closing pages his book, when he discusses how poor up-bringing and social conditions affect people's powers of reflective self-control: 1994: 231ff.) In addition, Wallace freely concedes that he 'largely takes for granted the notion of a general power' (1994: 192). Even if one agrees with Wallace that the metaphysics of rational agency should not form our starting point in thinking about responsibility, there is clearly more to be said about our moral and rational abilities.

5 Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (1998)

When Wallace's book appeared, John Martin Fischer had already begun developing an account of moral responsibility that emphasises our ability to engage in moral reasoning (Fischer 1982, 1987, 1994). The account is most comprehensively stated, however, in Fischer and Ravizza (1998). The book has been much discussed, and is included here as the most developed 'responsiveness to reasons' approach. I will suggest that Fischer and

---

people capable of participating constructively with each other in the exchange of moral criticism and justification' (1994: 234; see also 69).
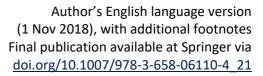
Ravizza's theory raises important issues, particularly in emphasising the importance of our *taking responsibility* for ourselves. While their account has important similarities with Wallace's, it also addresses the challenge that Wallace left open, the problem of how to understand our powers of moral reasoning. I will suggest, however, that their framing of the issues is not as helpful as it might be.

The basic outlines of the theory are as follows. What sort of control do responsible persons have over their actions? It seems unlikely, as many compatibilists have pointed out, that we are able to somehow choose between alternate possibilities, at least in a strong metaphysical sense. However, on occasion, we certainly deliberate between different possible actions, considering reasons for and against those options. In other words, we guide our actions in the light of reasons. Fischer and Ravizza argue that this sort of 'guidance control' is all that is needed for moral responsibility. Because this control is a familiar feature of our experience, moreover, we can leave open scientific questions about its causal basis and whether this should be interpreted deterministically. (Fischer and Ravizza call their account 'semi-compatibilist,' because it neither assumes nor denies the truth of determinism.)

Their main argument that moral responsibility does not require 'regulative control' over 'alternate possibilities' is that there are scenarios where a person is bound to act in a particular way but is nonetheless responsible. It might be that if a person does not choose to do so, a manipulator will cause them so to act. Harry Frankfurt (again) is the principal architect of such fantasies – for example, a mad scientist who can interfere in the firing of a person's neurones or some such (Frankfurt 1969). More sensibly, we can imagine situations where, unbeknownst to the actor, the possibility of acting differently is foreclosed by quite mundane, material aspects of her situation. Fischer and Ravizza give the example of a person who steers her car to the right, not realising that that other factors would prevent the car's doing anything else (1998: 32; also Fischer 2012: 39). While the example is artificial, there is clearly a sense in which the driver guides the car's

movement, and is responsible for turning right, despite the fact that she could not have made the car do anything else. This method of argument – which one might label "intuition pump by fanciful example" – is clearly not dispositive, since both examples and intuitions are open to question. So it should be added that this argument is not essential to the theory, inasmuch as Fischer (2012: 123) acknowledges the validity of separate arguments against the relevance of regulative control, such as those found in Strawson (1962).[10]

In any event, it is a separate matter to argue that 'guidance control' suffices for moral responsibility. Here Fischer and Ravizza point out that different causal pathways may stand behind actions. Clearly some pathways would undermine responsibility: if my action results from hypnotic suggestion, or my drink having been spiked, or the ludicrous machinations of an evil neuroscientist, then it was not my responsibility. The positive core of the theory, then, is that a particular sort of causal pathway does allow for moral responsibility – viz. that the action should issue from a 'mechanism' that is, first, responsive to reasons[11] and second, the agent's own.

---

[10]   However, we should note that first sort of example does seem to play a role in the structure of the theory, in particular by the place it grants to sub-personal 'mechanisms' in generating actions. I return to this point shortly.

[11]   This way of speaking has the advantage of allowing that much of our rational activity is not self-conscious. In common with all the authors discussed here, Fischer and Ravizza emphasise that we need not deliberate about reasons in order to think and act in ways that are rational: we need not, for example, think in terms of honesty or reasons to be honest in order to act honestly on a given occasion (thus their reference to 'nonreflective mechanisms of various kinds' (1998: 86), such as well-formed habits). However, we might note some reasons for caution. Especially when we focus on active deliberation, "responding" is a curiously passive term, while "responding to reasons" makes it sound as if reasons are entities of some sort. But responsible persons are *agents who reason*: reasoning is something that we do, well or badly, implicitly or explicitly. When responsible persons act and judge, they take certain considerations to count as reasons (to use Hieronymi's helpful formulation: 2011). In this respect, the language of judgment is happier than that of reasons: it is possible to hypostatise reasons, so that they seem to float free of an agent who
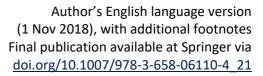
Both of these conditions are important, and neither hinges on the truth or falsity of determinism or the particular causal pathways underlying mental processes or bodily movements. With regard to the first condition, Fischer and Ravizza emphasise that the mechanism, such as 'ordinary practical reasoning' or 'deliberation involving an irresistible desire' (1998: 48), need not be totally reliable. After all, if we always responded properly to reasons, there would be no need to hold anyone responsible. But it must have a moderate degree of 'robustness.' A responsible person is not one who only responds to reasons in a few quite specific ways or exceptional situations. Evidently, this corresponds very closely to Wallace's talk of powers to grasp and apply reasons: in the responsible person, this ability is neither infallible nor sporadic.

But it is not enough that there merely be a moderately reliable 'mechanism.' Fischer and Ravizza imagine cases of persons who respond fairly reliably to reasons as a result of the machinations of a (non-evil?) neuroscientist or brainwashing or hypnosis.[12] Granting the method of "intuition pump by fanciful example," it is fairly clear that we should not see the manipulated person as herself a responsible agent. Both responsibility and reasoning rest with the person who pulls the strings. The mechanism that responds to reasons must be the agent's own.

What is it that makes a mechanism belong to a person? For Fischer and Ravizza, this requires a developmental process by which a person comes to see himself (i) as the source of his choices and actions and (ii) as responsible for them – for example, by

---

reasons. But no one – not even those who believe in God's final judgment – has ever supposed that there could be judgment without an agent who judges.

[12] As with many of their thought-experiments, one might worry whether such cases are coherent, never mind realistic, and thus whether the worries that their theory responds to are apt.

accepting that others may reasonably resent him if he acts wrongly toward them.[13] Most likely, this will be a matter of tacit acceptance, as a person grows up and is gradually acculturated in a community of responsible persons. By this process, a person 'takes responsibility for the springs of his action [and] makes them *his own* in an important sense' (1998: 210). They contend that failing to see oneself as the source of one's actions would mean that a person 'will presumably make no effort to affect the world by [her] choices or bodily movements' (218). And failing to see oneself as responsible would be to cut oneself off from normal human relationships, and even to lose the unity of the self. They give the example of a serial killer who blames "a monster inside" (a non-reasons-responsive mechanism if ever there were one!) for his crimes (219f).[14]

Fischer and Ravizza clearly identify important factors here. Invoking different 'mechanisms' has the advantage of allowing that one and the same person may sometimes be responsible for her deeds and sometimes not. She is responsible when her actions issue from the mechanism of 'ordinary practical reasoning' (216); she is not, when her deeds issue from a mechanism impervious to reason, perhaps one involving a compulsion or addiction or derangement. Further, the idea that responsibility arises as we *take responsibility* for our agency is a pregnant one, as is the idea that taking responsibility is essential to full participation in normal human relations. It is, I think, a shame that they develop these latter points only briefly, and I will offer a few further thoughts in a moment.

It is also worth noting some difficulties posed by their framing of the issues. In the first place, I have already noted that interpreting the human capacity for reasoning as a capacity for 'responding to reasons' can sound overly passive. To figure this capacity as a

---

13   They further specify that these beliefs must be appropriately based on evidence, in order to address problems posed by outlandish forms of manipulation that involve 'implanting' such beliefs.

14   While both points are suggestive, I suspect their formulations are over-hasty – I give them here just to convey the overall line of thought.
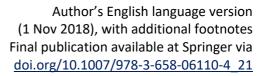
'mechanism' is a much more radical step, however. It reframes *something we do* as something that *happens* – either within us or, in the more bizarre manipulation cases, outside of us.[15] Putting things this way opens a sceptical gap: how can this happening be a person's doing? (In passing, it is worth noting that this same sceptical gap often motivates incompatibilist views. If we look at human activity quasi-mechanically, in terms of causal processes, then it is hard to locate an agent who does anything.) Applied to reasoning, mechanistic images are obviously out of place: there is no reasoning without an agent who reasons.[16] In other words, the idea of 'taking responsibility' is invoked to solve an artificial problem. Moreover, it creates a circularity: someone's mental activity – presumably a process of reasoning, however inarticulate or tacit, and hence on Fischer and Ravizza's terms, the activity of a mechanism – is invoked to explain how the activity of a mechanism represents someone's responsible action.

This said, these perplexities are not simply artificial, even if they are exacerbated by an ill-considered 'mechanism' metaphor and a preoccupation with imaginary examples. Underlying them is a very fundamental problem: just how are we to understand rational and responsible agency, and the relevant notion of ability?

I will return to questions of agency and ability shortly, in considering the work of Angela Smith and Pamela Hieronymi. First, however, it is worth spending a moment more on the notion of 'taking responsibility.' While Fischer and Ravizza's framing of the issues may be forced, the circularity just highlighted is arguably inherent in our subject matter.

---

[15] Fischer himself denies that talk of mechanisms is meant to be mechanistic (Fischer et al 2007: 78). I also note in passing a further problem that lurks here. As several critics have wondered, how we are to decide what sorts of things should count as a mechanism, and whether *the same* mechanism at work in someone's thoughts or actions? (McKenna 2001; Levy 2007) Beyond a characteristic appeal to intuitions, this is a point that Fischer has never clearly responded to (see Fischer 2006: 239ff).

[16] If this claim sounds too strong, perhaps because you suppose that (say) a chess computer reasons before it makes a move, then the point can be put in a more restricted way: There is no *moral* reasoning without an agent who reasons.
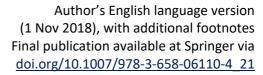
None of us made our own agency; each of us is formed by social and biological factors that necessarily lay outside of our control; each of us is finite and fallible. Yet, if morality and rationality are to be effective forces in the world, each of us must strive to incorporate these demands within our agency.

In part, taking such responsibility depends on a commitment to live up to valid norms. In this case, others will not have just cause to hold you responsible because you have already taken that responsibility within yourself. But taking responsibility also has a retrospective element: a readiness to hold yourself responsible, a willingness to accept and learn from valid reproaches. To mean anything, these must involve a reluctance to cavil and pass the buck, or (most importantly for our theoretical debates) a disinclination to whine that you could not have done otherwise because this is how your biology and parents and society made you. However we understand the history by which we came to be who we are, we may not simply disavow our activity as the mere product of causal forces. As Wallace points out, this abuses the very idea of an excuse. Excuses indicate that, despite appearances, we are not committed to a particular, wrongful way of thinking or acting – not that we are incapable of bearing any commitments of our own.[17] And as Robert Adams pointed out in a famous article, 'To refuse to take responsibility for one's emotions and motives is [first] to be inappropriately alienated from one's own emotional and appetitive faculties,' and second, to refuse a crucial opportunity for moral learning (1985: 16).[18]

---

[17]   Here I am looking at matters in bare outlines only. Clearly there are many complexities concerning, for example, the place of self-forgiveness and self-understanding, or how we should understand (and feel) guilt.

[18]   Adams' thought is worth quoting at more length: 'Your ingratitude… is not a voluntary action; but if you take responsibility for it you also do not see it as something that just happens to you, like a toothache or a leak in your roof. You see it as an opposition that you yourself are making, not voluntarily but none the less really, to the generosity of the other person and to your own position as a recipient of love and assistance. In repentance you
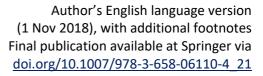
In other words, taking responsibility is fundamental to responsible agency.[19] I already suggested, in response to worries about the concept of ability arising in Wolf's account, that there might be a constitutive link between responsible agency and practices of holding responsible. The key point to notice, now, is that norms telling us to take responsibility for ourselves – to hold ourselves responsible, as one might also say – help to constitute responsible agency, and thereby to render meaningful our practices of mutual accountability. It is not only a sociological generalisation that people tend to 'take responsibility' for their actions and selves as they grow up. It is also an essential duty: we must learn to understand our own activity and evaluate it in the light of normative requirements, and to accept that other people may sometimes be able and entitled to correct our evaluative stances. Fischer and Ravizza can plausibly contend that someone who fails to take responsibility for himself *tout court* simply lacks the capacity to 'respond to reasons,' and hence is non-responsible. But behind this claim stands a more fundamental point. Whatever else we say about the reasons that responsible persons

---

repudiate this opposition, not as an evil existing outside the inner circle of your selfhood, but as your own; and you reproach yourself for it.' (1985: 15f)

[19]   As mentioned near the start of this chapter, following Frankfurt 1971, another strand in the responsibility literature suggests that free action is action on desires or attitudes that a person 'identifies with.' It may be worth commenting on how this differs from a view that emphasises 'taking responsibility.' As I pointed out above, the identification strategy is morally out of kilter. The person who acts on a desire or attitude that she disavows or disapproves of is not, thereby, relieved of responsibility: self-deception and weak will are often blameworthy. The 'taking responsibility' strategy, if I may put it that way, agrees that a person who takes responsibility for her attitudes or actions must 'identify with' them. But this is only in the sense of accepting that particular attitudes or actions were manifestations of her own agency. Hence 'taking responsibility' is most clearly visible when we feel regret and perhaps guilt: I should not have felt that, thought that, done that... The person 'identifies with' her action (etc) in the sense that she acknowledges it as an exercise of *her* agency; and she disavows it, in the sense that she asks forgiveness and resolves to do better in the future. This is not to deny, of course, that a person may just as well 'take responsibility' for actions that she sees as neutral or praiseworthy.

are meant to 'respond to,' they must include reasons to take responsibility for oneself, to understand oneself as the active and accountable bearer of commitments.

6   Angela M Smith: responsibility goes through judgment, not choice

In the last two sections of this chapter, I consider the work of two authors whose positions overlap in significant respects. Angela Smith and Pamela Hieronymi have both pursued one of the central conundrums of rational and responsible agency. The puzzle goes back to the basic problem facing compatibilists and incompatibilists alike. How are we to situate agency, such that we can understand its relation to biological and social processes, such that it gives rise to empirically observable events that belong to a causally well-ordered world, yet such that it stands at sufficient distance from these to represent an independent, accountable source of activity? Or in terms of the perplexity posed by Fischer and Ravizza's mechanical metaphor: how can we make a place for doings in a world of happenings? Incompatibilist answers tend to make this agency mysterious: the famed 'ability to do otherwise,' for example, points to a problem without solving it. Yet the compatibilist approaches discussed above fall back on the idea of rational ability – a power to think and judge and act in the light of reasons. This ability may be practically familiar. But it still requires philosophical elucidation.

One way of gaining some orientation in these deep waters is to pay closer attention to the relation between responsibility and control. Following Adams (1985), several recent authors have pointed out a simple problem with the common thought that 'persons are responsible for what is under their control': that it misses many significant attributions of responsibility. For instance, George Sher's article 'Out of Control' (2006) memorably runs through examples of wrongful action due to neglect, inattention, carelessness, panic, hasty judgment, insensitivity and moral mistake. In none of his examples does it seem right to say that the person controlled what she did (or failed to do), in the sense
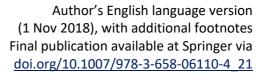
of having wittingly chosen to perform a wrongful action. Yet in each case, it is fairly clear that the person deserves to be blamed for her action or omission.

Similarly, Angela Smith centres her article 'Responsibility for Attitudes' (2005) on the problem of attitudes that reveal themselves in (for example) forgetfulness, spontaneous reactions, thoughts that merely 'occur to us,' and the things that we 'happen to' notice. (That we apply the language of happening and occurring to our mental life already suggests an interesting combination of activity and passivity.) Smith's main claim is that the kind of activity for which we hold one another – and ourselves – responsible is 'not the activity of choice, but the activity of evaluative judgment' (2005: 237). Spontaneous reactions and thoughtless omissions, no less than deliberate actions, reflect on the actor insofar as they enact her sense of what matters.[20] People do not and need not actively formulate this evaluative sense, and yet it would not be right to describe it as something passive. It characterises a person, but it does not merely *befall* her.

To avoid misunderstanding: Smith is well aware that some mental states are essentially passive. Under normal conditions, I am simply subject to sensations such as pain or thirst. These do not yet involve judgment or evaluation, and they do not pose questions of responsibility. (Questions of responsibility may arise by virtue of the attitudes I take toward these sensations, however: for example, if I were to neglect the bodily needs they reflect, or alternatively, to neglect a more urgent matter.) Sensations have, at most, a causal connection with our attitudes: for instance, an attitude of fear – which belongs to a judgment that something poses risk or danger – might cause a mere sensation such as nausea; conversely, if I fail to take care of myself, then thirst and hunger are predictable results. Moreover, sensations can only be useful or damaging, pleasant or

---

[20] Adams (1985) makes the same point with his opening example: I may be able to control the expression of my anger, but I cannot choose whether to be angry or not.

unpleasant. By contrast, judgments are subject to standards of right and wrong, reasonable or irrational. So too are the attitudes that instantiate those judgments.

Smith's basic thought, then, is that our thoughts and attitudes, acts and omissions reveal where we stand on all sorts of normative questions – on what is owed to friends as opposed to strangers, for instance, or the priority of different projects and concerns.[21] If I forget a friend's birthday, to use Smith's example, this shows something about the event's importance in my scheme of things (that it is less important than a family crisis or my latest love affair, as the case may be). Evidently, this attitude may be faulty; but what is at fault is a pattern of attention and commitment, rather than a choice. Smith's point is that this pattern reveals my priorities and enacts my judgments (where judgment should be broadly construed, as an evaluation of the world and my place within it, rather than as a specific cognition at a particular point in time [2005: 251f]). And there is, in a sense, nothing more important about a person than these evaluations. Thus an act of forgetting may be blameworthy, or at least call for justification, because it has, as Smith puts it, a 'rational relation' to an evaluative judgment – to my sense of what matters. Smith sometimes refers to 'judgment-sensitive attitudes,'[22] but her point is not that our attitudes are *caused* by our judgments.[23] In a fully rational agent, judgments would correspond to attitudes, responses and the rest. In a partly rational agent, such as we all are, attitudes (etc) *should* correspond to judgments (2005: 253).[24] To use Smith's examples again: If I believe that household spiders are harmless, I should not react to them

---

[21]   As Smith notes (2005: 242), unconsidered actions and responses may often be more revealing than deliberately chosen deeds: we betray ourselves (for better and for worse), rather than presenting what others expect of us.

[22]   Following Scanlon 1998: 18–22.

[23]   As Hieronymi points out, to see things in this way would place our attitudes 'at an odd distance from our reasoning' (2009: 214); see similarly Smith (2015: 127).

[24]   For this reason, Hieronymi speaks of 'commitment-constituted attitudes' (e.g. 2008: 361f, n10)
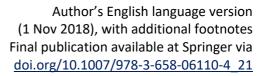
with fear; if I care about my friend, then (barring crises and emergencies) I should remember her birthday. As a result, a person may be open to praise or criticism on several dimensions: whether her attitudes-as-manifest-in-action are reasonable, whether her underlying evaluative judgments are reasonable, and whether her actions and attitudes and judgments cohere with one another.[25]

Smith's deeper point concerns how we construe the activity of responsible persons. She contends that 'the dividing line between activity and passivity goes through judgment' (263). The distinctive power of a rational agent is not the ability to act as an uncaused cause; it is the ability to stake an evaluative position and to give and ask for reasons to justify such positions.[26] And in everyday life, that giving and asking is usually not a matter of polite debate around a seminar table – it may well extend to angry accusations that an action or omission or attitude is out of order. By extension, the metaphysics we need to interpret responsible agency should not focus on questions of causation or determinism, but rather on the distinctive nature of rationality and evaluative judgment. In particular, action must appear, not as an event caused in a particular ways, but rather as the manifestation of a rational power: the power to judge what matters and to act in the light of this.

Like all authors in these debates, Smith is well aware that upbringing and cultural context profoundly shape our evaluative commitments and our abilities to revise them. She acknowledges, then, that further questions can arise about a person's responsibility for becoming who she is (2005: 267f, 2008: 389f). Answers to these questions may well

---

[25] At the same time, these dimensions are not wholly distinct. As Smith points out (2005: 255), without a certain degree of coherence, it is hard to attribute a definite mental state to a person. In cases of moderate incoherence, we may speak of ambivalence or weakness of will, or more figuratively, of wanting to have one's cake and eat it.

[26] Although we may revise our attitudes, in the light of our other attitudes and our sensitivity to different considerations, we do not produce them. As Smith says, 'we are, first and foremost, inhabiters of' our attitudes (2005: 251) – and not less active for that.

affect how we respond to someone.[27] But they do not alter the fact that a person enacts certain evaluative commitments as she moves through the world; nor do they touch the fact that these commitments are vitally important (whether for better or for worse) to the people around her; nor do they change the fact that those commitments ought to be rationally defensible. Moreover, whatever allowances we are inclined to make when we understand how someone came to be who she is, we should be reluctant to view people as 'passive victims of their faulty judgments' (2008: 390), or indeed, of their upbringing and circumstances. Such a stance may be seen as patronizing or disrespectful. One might add: it is also peculiarly hard to maintain when face to face with someone who sees no reason to alter her unreasonable or immoral conduct.

As mentioned, the most obvious issue at stake here is responsibility without control. It is highly implausible to hold, for example, that spontaneous, unconsidered reactions are morally insignificant and do not reveal morally important character traits. I am blameable for the glee I feel at a friend's misfortune,[28] and even more blameable if I betray that glee, however involuntarily. Some philosophers who advocate the importance of control for responsibility would argue that, at some previous point, there must have been an exercise of my choice.[29] But if we consider the idea a person may have actively decided to become insensitive or spiteful, or is only culpable for her spontaneous insensitivity and spite if she decided to become such a person, we can see that such arguments are

---

[27]   Or indeed, the meaning we attribute to her action, and hence the underlying commitment – thus Cheshire Calhoun (1992) and Kyla Ebels-Duggan (2013).

[28]   Thus Gore Vidal's famous *mal mot*: 'Every time a friend succeeds, something inside me dies' (used as the title of Jay Parini's 2016 biography).

[29]   Neil Levy (2005), for example, takes such a position, by distinguishing between 'bad' and 'blameworthy' agents (see the discussion below, in section 7, on Watson's distinction between 'attribution' ['bad'] and 'accountability ['blameworthy']). See also Otsuka 1998, Michael Smith 2004.

procrustean.[30] They are driven by theoretical preconceptions rather than the realities of human agency and interaction.
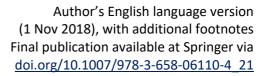
In any event, a deeper issue lies behind the implausibility of such attempts to rescue a control condition for responsibility – or alternatively, behind their *plausibility*, since they arguably reflect the difficulty of understanding how activity could be both *ours* and not under our control.[31] Just what sort of activity is involved in rational and responsible agency, and the often spontaneous judgments and actions that manifest it?

7 Pamela Hieronymi: responsible activity as 'evaluative control'

To consider this question, I now turn to the work of Pamela Hieronymi. Hieronymi's substantive views on responsibility are very similar to Angela Smith's, but she has focussed in more detail on the nature of mental activity. In essence, she argues that if we consider more carefully what is involved in this activity, we will see that – important as it is – control is not the primary face of responsible agency.

---

[30] As Smith (2015) argues, there are clearly cases where a person's previous decision is important to the blameworthiness of an involuntary action. For example, we may blame the driver for a car crash because she drank alcohol beforehand. This sort of involuntary action, however, is very different from spontaneous emotions or reactions or forgettings or failures to notice. Some involuntary actions and feelings straightforwardly reveal who we are and what we care about; there is no need to look for some prior act of control, because we just *are* the person who thinks and feels in these terms. By contrast, the drunken crashing of a car is involuntary in a different sense: our agency has broken down and no longer reveals what we think or care about. We must therefore look to previous actions or attitudes (or subsequent responses) in order to assess the person.

[31] Here, one of Hieronymi's diagnoses may be helpful. She points out that we may regard voluntariness as a criterion for responsibility because only voluntary actions reveal our evaluative commitments. However: 'These attitudes need not be voluntary in order to reveal one's moral personality. Rather, they constitute one's moral personality – moreover, they do so precisely because they represent one's take on the world, and so could not be voluntary' (2008: 372).
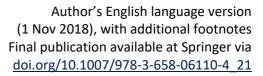
Her main argument is as follows. Large swathes of our mental lives are not, and could not be, subject to our control, in the following sense.[32] We cannot choose whether to believe something, or resent something, or intend something, on the basis of any old reasons that might make it appealing to maintain that mental state.[33] To stay with the example of belief: I can only properly believe something for reasons that count in favour of its truth. It might, perhaps, make me happier to believe that I am more popular than I am. But that is not the sort of reason that could or should convince me that I am

---

[32] A similar line of thought is also suggested in some of Smith's work (e.g. 2004; 2005: 264f). In addition, Richard Moran (e.g. 2012) and Matthew Boyle (2011; 2013) make parallel arguments about the nature of agency: that it is not at root a matter of control, in the sense of being able to alter matters at will. As Moran puts it, 'I do not aim at acquiring some particular attitude, and its rationality is not expressed by my singling it out for control or manipulation. But for all that, my wanting, suspecting, and caring about something are expressions of my active nature, to which some form of the normative 'why?' question naturally applies, along with my taking myself to be the person who is answerable for why I do the things or believe the things that I do' (2012: 219). To this I would add: it is another matter, how well a person is equipped to answer those 'why?' questions, or to recognise that his answers to some of those questions are inept or inapt.

[33] The example of intention is something of an outlier here, inasmuch as intentional deeds seem to represent a paradigm case of control. To see Hieronymi's point, we need to consider cases where I have reasons to intend to do something, without having reasons to perform the action. Examples of such cases tend to be more contrived (they have their original in Kavka 1983). False beliefs can obviously provide psychological payoffs for the believer. Cases where separate reasons bear on the having of an intention than on actually performing the intended action arise only when we can drive a wedge between intention and deed – and those two things, of course, have very close conceptual and practical connections. However, a gap can open up if other people impose inconsistent or irrational demands: in particular, if they care, for whatever reasons, much more about your attitudes than the actions that eventually result. A lover who is ambivalent about marriage, for example, may sincerely desire that you intend to marry him, without wanting you to actually do so; a boss who is concerned more about her power position than your performance may care more about your intention to perform various tasks than whether you finally do them (see Hieronymi 2006: 57, 2008: 368f). Such cases present you with an extrinsic reason to intend, a reason to adopt an intention which is not at the same time a reason to perform the intended action. In such cases, a lack of voluntary control again shows up in the fact that, however desirable it would be to have the intention without its intrinsic connection to action, this is not actually tenable. As belief hinges on what is the case, intentions hinge on what we have reason to do.
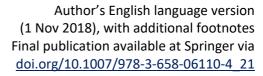
especially well-liked. (If I were somehow to convince myself of my popularity, in the teeth of the facts, that would be mere wishful thinking.) As is often said: beliefs aim at truth. As we might say: believing is responsible to the facts. But although we cannot believe "at will," believing is nonetheless something that we do.[34] And it is something that we are answerable for: we should be able to give some account of why we maintain a belief, and others may reasonably criticise us for ill-founded beliefs – for example, if we engage in wishful thinking.[35]

As indicated, belief is just one example. We exercise, as Hieronymi puts it, 'evaluative control' throughout our mental activity – in the intentions we adopt, the gratitude or resentment or good will that we feel toward others. These are all things that we cannot do "at will." Instead, they reflect our 'take' on the world and what is worthwhile or worth doing within it.[36] Without this background, no choice could appear sensible or

---

[34] This claim may be controversial, but is certainly at the heart of the wider philosophical issues at stake here. There has, of course, been a tendency in modern philosophy – exemplified by Locke's epistemology and memorably parodied by Laurence Sterne in *Tristram Shandy* – to view beliefs (among other attitudes) 'merely as dispositional states that we might find ourselves with' (Hieronymi 2010: 9). Leaving aside the very large question of whether such a view is tenable, let me just note the significance of the point. To see beliefs (etc) in this way is to pose the challenge of how they might be incorporated into a person's activity (or alternatively: how they might count as *beliefs*, if they are not already so incorporated). This is the theoretical analogue of a more practical point made by Adams (quoted above at note 18), that such a view involves an 'inappropriately alienated' attitude toward our own mental lives; in more moralistic terms, it represents a failure to take responsibility. See also Korsgaard 2010.

[35] This point is sometimes obscured when authors claim that believing, unlike acting, does not raise questions of *moral* responsibility (e.g. Wallace 1994: 80f; Darwall 2006: 10ff). But this is not true. Some cognitive mistakes may be morally innocent, inasmuch as they do not reflect on our attitudes to other people (Smith 2008: 386f). But in all practical contexts we have duties to take account of many facts: failures to assess these adequately may be extremely serious, and many vices are bound up with characteristic cognitive distortions. Consider, for example, the combination of ideological bias and self-servingness bound up with denial of anthropocentric climate change.

[36] To spell out Hieronymi's thought a little more: We cannot rearrange the contents of our own mind in the same way that we can sometimes rearrange our bodies or environments.
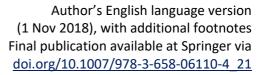
reasonable to us. Our deliberate actions reveal this evaluative activity, no less than our spontaneous responses; it is often implicit, too, in what we omit or forget.[37]

There is a second respect in which my 'control' is limited.[38] I cannot simply choose to get things right. With regard to belief, for example, I can draw on more-or-less reliable methods to check certain beliefs, and keep up my guard against known sources of error. But these methods offer no guarantees and I can only employ them sporadically. Given that 'judgment' is a theme of this chapter, we might also note how profoundly this point applies to our assessments of persons and situations. Such judgments are bound up with our emotions and habits in all sorts of ways, and together these make an enormous difference to how we respond to others. At the same time, they are profoundly fallible. Although people may devote more or less effort to examining (for example) their biases or assumptions or emotional habits, they obviously cannot step outside of themselves to revise their interpretations of people and the world and their place within it. A person cannot leave behind the frame of his own mind. Such 'control' as we have can only be exercised by means of the abilities we already possess and from within the perspective we already occupy.

Or rather: sometimes we can, but this involves forms of self-manipulation that rely on the fact that our minds work in more or less predictable ways. Someone prone to wishful thinking may convince himself that something is the case; he might even take quite deliberate steps – for example, by watching rom-coms and avoiding accurate news reporting, he might kid himself that all is well with the world. Or someone aware of his tendency to see things through rose-tinted glasses might aim to correct this – for example, by working his way through studies of political evil. The idea that this sort of control could be the usual case, however, plainly mistakes the nature of agency. Such manipulations depend on the existence of reliable patterns of thought and feeling, which amounts to saying that our mental activity is *not* subject to our control.

[37] By mentioning forgetfulness, I mean to remind the reader that this is also Smith's position.

[38] Here I depart from Hieronymi's own framing of the issues, in order to make more vivid the connections between our fallibility, questions of control, and practices of mutual accountability.
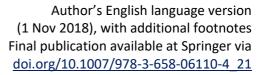
That last claim is too hasty in one vital respect. We share the world with other people and we may, if we are fortunate, learn from their perspectives and benefit from their abilities to judge and reason.[39] This should remind us of an issue that arose with regard to Wolf's notion of ability: the ability to revise our judgments can be augmented or undermined by social context. If we are (morally) fortunate, others will point out our failures to care about certain aspects of life. In particular, they may do so by blaming us for our failures. That is to say, practices of holding people responsible can help us to 'see' things that we 'could not' see before, helping us to take responsibility for defects in our judgments and reasoning. Advocates of a control condition for moral responsibility may argue that a person should only be judged *responsible* for such failures if she 'could have' thought or acted better on her own initiative, or (more modestly) if she can appreciate the fault and alter her activity once it is pointed out to her. Hieronymi, among others, points out some crucial problems with using ability as a criterion for responsibility attributions.[40]

In the first place, it would be very strange for a normative standard to retreat in the face of inability. If I am not able to do maths well, that does not mean my faulty calculations should be accepted or endorsed. If I am not able to look after a child properly, that means I should not act as a parent. Why should matters be any different if I prove unable to meet normal moral standards – for instance, if I am consistently unreliable or aggressive? These are serious faults with serious consequences for those around me. From others' points of view, they surely remain faults even if (perhaps: especially if) they are cannot be corrected. Other people have every reason to criticise and indeed resent my failures to respect them. They have every reason to alter how they relate to me and to

---

[39]   Clearly, vital questions of judgment arise here. We may be too ready to be swayed by others, or by the wrong people; or we may not be open enough – stubborn or insensitive or closed-minded.

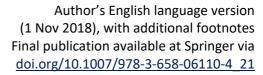[40]   A cognate argument is made by Smith 2008 and Talbert 2012.

avoid certain sorts of relationship with me. In other words, they have good reason to treat me as responsible. As Hieronymi says, 'moral demand need not recede in the face of an individual's inability to meet it. Rather, legitimate moral demand requires only that one have the more general capacity to stand in interpersonal relationships' (2007: 122).

Opponents of this position argue that the ability to do better is required on grounds of fairness, as mentioned in our discussion of Wallace's account. If someone is unable to live up to a moral standard, this means she lacks a 'fair opportunity' to avoid the 'sanction' of being held responsible. In his well-known paper, 'Two Faces of Responsibility' (2004), Gary Watson tried to accommodate this point by separating judgments of 'attribution' from those of 'accountability.' He allows that we may attribute virtues and vices to people independently of questions of control. The incorrigibly selfish person is indeed selfish: her thoughts and actions are governed by an unmistakable pattern of commitments and evaluations. Such moral appraisals are apt insofar as the person has *some* normative competence; but this competence need not extend to an 'ability to acquire the right values' (2004: 282). When it comes to matters of accountability, however, fairness dictates that people should have the opportunity to avoid adverse consequences such as retribution or demands for compensation, so that questions of voluntariness and control remain relevant. As Watson acknowledges, however, we need to say more than this. Social regulation also matters: for the sake of other people's rights or standing, we may have to impose costs or limits on the incorrigibly vicious person. Watson attempts to maintain his opening distinction by claiming that such responses then 'lose their normal expressive function' (2004: 281).
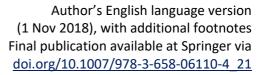
This position is clearly uncomfortable. The incorrigibly vicious person is to be appraised as vicious *and* to be restrained in the exercise of her vices for the sake of others. The difficulty is to see how such treatment differs from that we should apportion to someone who we believe to be (so to speak) corrigibly vicious, that is, who has the ability not to *be* vicious even though her actions *are* vicious. (We might also wonder how we should

tell the difference between these two characters…) There are, I think, two related problems here. First, as Smith argues, Watson seems to have underspecified the sorts of responses that we make to wrongdoers (2008: 375ff; 2012a). Second, as Hieronymi (2004) points out, this neglect reveals a basic problem with the 'fairness' argument itself. If we consider the specific actions involved in holding a person responsible, it is simpleminded to frame these as costs or burdens on the person (at least so long as we set aside deliberately punitive responses). Consider, for examples, blame or feelings of resentment or tendencies to alter the terms of our relationships with someone. Blame rests on a judgment that actions and character traits are faulty; insofar as our self-respect is embodied in characteristic emotional responses, we will be liable to feel resentment when someone infringes upon it; it is obviously a mistake to rely on someone who is unreliable or to allow free play to another's aggression. Such judgments and responses are apt by virtue of how a person treats us. By the same token, they do not become more or less apt by virtue of further considerations about whether they add to the weal or woe of the person who abuses us, or whether these effects would be 'fair' in the larger scheme of things. Just as it would be wrong to believe something simply because it would make me happy, these would be the wrong sorts of reason to withdraw (or award) criticism and resentment.

This said, there are certainly important practical questions about when we should voice our blame, how we should channel our resentment, the ways in which we should modulate our relations with other people, and so on. The ways in which another person is able to, or likely to, respond to our responses will often matter when we answer such questions. But then it is important to remind ourselves, again, how complex our practices of holding responsible are, and how they must interact with the many other sorts of reasons governing how we should treat one another (see especially Smith 2007). In other words, the apparently simple question 'Is a person responsible for her vices or misdeeds?' is not enough to guide us here. If we set aside all the other reasons that we
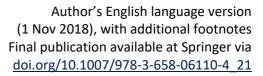
have to treat one another in particular ways, we end up stretching the principle of responsibility past breaking point. It should hardly be a surprise that it ends up seeming unfair, or otherwise dubious.

In summary: A person's control over his activity is limited in two vital respects. (i) Judgments and attitudes are not and cannot be subject to our control, at least not in the normal case: we cannot believe or intend or resent or fear or love "at will." Rather, we should understand these attitudes as instantiating our rational and evaluative activity. Another way of putting this is to say that the control we exercise lies 'behind the lens' (Hieronymi 2008: 371). We control our thoughts by thinking them; more broadly, we control our evaluative activity by judging, reasoning, gradually learning what is fulfilling to ourselves and esteemed by others.[41] If we think of responsible agency as a whole, self-conscious exercises of control, such as deliberate actions or attempts to remould particular attitudes, are more like the tip of the iceberg than the foundational or paradigm case. (ii) As Fischer and Ravizza insist, we are only 'moderately' responsive to reasons. It would be, in fact, very odd to justify practices of responsibility on the basis that we are *able to get things right*. Whatever else they are, our abilities to reason and judge are *fallible*. To suppose that failings could be subject to our control again gets things back to front. No doubt there are wilful failures and deliberate wrongs, but these are the exception rather than the rule. Behind them lie patterns of attention, broader evaluative frameworks, tendencies to take some things and people more seriously than others. These may, in the worse cases, insulate a person from appreciating reasons to think and act differently; even in the best cases, they can blind a person to specific reasons or sources of value. The particular limitations on each person's agency give others more,

---

[41] To quote the passage in full: 'We change our minds, and so control our attitudes, not by reflecting on or thinking about our mind, but rather by thinking about the object of our thoughts. The controlling happens "behind the lens," so to speak. The thinking subject controls its thoughts in thinking them.' (Hieronymi 2008: 371)

not less, reason to scrutinise his actions and evaluations – at least until such point as they abandon *all* faith in his ability to judge and reason about the ways we should share the world with one another.

For these reasons, both Smith and Hieronymi emphasise not choice but judgment. Our capacities for deliberate action and self-control are necessarily limited. Yet as she moves through the world a person cannot but take a stand on what matters and the terms on which she lives with others. Regardless of whether someone explicitly (re)considers these evaluations, they will be implicit in her thoughts and deeds, and of profound consequence for those who interact with her. To understand this, we need, first, an account of (responsible) activity that does not hinge on the idea of a separate controller and, second (though this is perhaps only hinted at in the work of the authors considered here), an account of mutual recognition, whereby responsible persons understand one another – and themselves – as the authors of such activity and the bearers of corresponding commitments.[42]

---

[42] Here I extend (or bend) slightly Hieronymi's own remarks in the last section of a recent paper, when she stresses the importance of 'a certain kind of interpersonal recognition' to responsible agency (2014: 40). My own suggestion would be that it is essential to normal human relations that our thoughts and actions bear a special significance for other people: not only as damaging or pleasant, like hailstones or sunshine, but as instantiations of evaluative judgments. To act as the bearer of such judgments requires, of course, many individual abilities, some of which may be absent or ill-formed in the vicious, just as they may be well-formed in the virtuous. But it also depends on the stance that others take towards us, when they understand us as responsible persons whose attitudes and activity count as contributions (valid or faulty as the case may be) toward a normatively ordered world. This is, I think, the deeper truth underlying Smith's point that it may be patronising and disrespectful no longer to measure a person against normal moral standards. Whatever we do to guard against the physical or social harms a person might cause, the truly radical step is to stop thinking of her as a fellow judge of how we should interact.
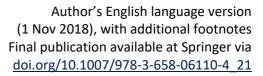
8 Conclusion

The contributions considered in this chapter all provide important help for understanding the relationship between responsibility, rationality and judgment. All the authors give us compelling reasons to connect responsibility and our capacities to reason and judge. And all of them give us reasons to think that these capacities are compatible with the causal order of the world, even if they require distinct normative categories to become visible to us.

I suspect that two large problems continue to motivate incompatibilist doubts. First, with the significant exception of punishment, there has been a severe lack of attention to actual practices of holding responsible: only over the last few years have these become a theme (not least, in contributions by Smith [2007, 2012b] and Hieronymi [2004]). A few authors are quite candid in holding that responsibility cannot be justified in compatibilist terms because they take their bearings from a notion of responsibility that human beings could never enact – thus Galen Strawson's notorious references to "heaven-and-hell responsibility."[43] More generally, as Peter Strawson pointed out (1962), if we focus on the abstract question of whether or not people are responsible, or may fairly be held responsible (where this is understood, simplistically, as a harm or burden), then we lose sight of the relationships and interactions involved in practices of responsibility. Losing sight of our subject matter, it should be no surprise that responsibility becomes more mysterious rather than less.

Second, as I have suggested at several points in this entry, the broader notions of rational agency and moral ability have been underexplored. For all the ink that has been

---

[43]   'As I understand it, true moral responsibility is responsibility of such a kind that, if we have it, then it *makes sense*, at least, to suppose that it could be just to punish some of us with (eternal) torment in hell and reward others with (eternal) bliss in heaven' (2008: 332). Put this way, the difficulty should be obvious: why should something so rebarbative answer to the 'true' version of anything – unless, perhaps, of an atavistic fantasy?

spilled in debates between compatibilists and advocates of metaphysical free will, strikingly little has been said about the notion of ability, and especially about abilities to judge and reason. J L Austin once commented, 'It is *can* in particular that we seem so often to uncover, just when we had thought some problem settled, grinning residually up at us like the frog at the bottom of the beer mug' (1961: 231). The Kierkegaardian frog, one may feel, is still there, grinning or grimacing up at us.

These two points intersect in interesting ways. As noted at a few places in this chapter, there are reasons to think that practices of holding responsible have a constitutive link with our abilities to act as – *to be* – rational and responsible persons. Without the company of decent persons who are willing and able to hold us responsible on decent terms, human capacities for rational thought and moral judgment often prove feeble. This is not to deny that people should be blamed for complicity in widespread evil, or otherwise held responsible for this. But it does suggest that (what look like) individual capacities may have a social dimension, and that practices of holding responsible help to realise this (Vargas 2013).

In addition, we might also speak of the sociability that responsible persons enact within themselves, when they consider other people's perspectives and possible responses to different modes of thought and action. Fischer and Ravizza rightly highlight each individual's ability-cum-duty to take responsibility for his own activity. As we might also put it, it is foundational to responsible agency that each of us learns to hold herself responsible. Or even more strongly: this belongs to rationality itself.[44] Without the ability to take others' perspectives within ourselves, and to admit that they may represent valuable correctives to our own activity, a person could never be fit to participate in social practices of holding responsible. These would become ways of merely exerting

---

[44]   In her work on self-constitution, Christine Korsgaard (2009) may be the philosopher who comes closest to this view, arguing that rational agents must learn to 'pull themselves together.'

pressure[45] – at best, the utilitarians' well-intentioned but manipulative economy of threats; at worst, demands to conform in oppressive or unjust social arrangements.

However we finally understand these matters, I hope this chapter has indicated the range of issues at stake. Practices of holding responsible hang together with the distinctive human capacities to reason and to act on the basis of moral considerations – fragile and fallible as these capacities are bound to be. This chapter has only touched in passing on the rationality of those practices. However, perhaps enough has been said to indicate that practices of holding responsible and taking responsibility are closely interrelated, and that both play an important role in sustaining rational agency and moral community.

9   References

Adams, Robert Merrihew. 1985. Involuntary Sins. *Philosophical Review* 94: 3–31.

Austin, J L. 1961. Ifs and Cans. In his *Philosophical Papers*, Oxford: Clarendon Press, 153-80.

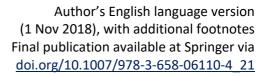Boyle, Matthew. 2011. 'Making Up Your Mind' and the Activity of Reason. *Philosophers' Imprint* 11 (17).

- 2013. Active Belief. *Canadian Journal of Philosophy* 39 (suppl. 1): 119–47.

Benson, Paul. 1996. Review of R Jay Wallace, *Responsibility and the Moral Sentiments*. *Journal of Philosophy* 93: 574–78.

Calhoun, Cheshire. 1992. Changing One's Heart. *Ethics* 103: 76–96.
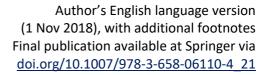
---

45   Bernard Williams makes this point, somewhat archly, in his well-known remarks about blame and the 'fiction' of deliberative community (1985: 192ff).

Coates, D Justin & Tognazzini, Neal A (eds). 2012. *Blame: Its Nature and Norms*. New York: Oxford University Press.

Cullity, Garrett. 1997. Review of R Jay Wallace, *Responsibility and the Moral Sentiments*. *Mind* 106: 803-807.

Darwall, Stephen. 2006. *The Second Person Standpoint: Morality, Respect, and Accountability*. Cambridge MA: Harvard University Press.

Dennett, Daniel. 1980. The milk of human intentionality. *Behavioral and Brain Sciences* 3: 427-430.

Ebels-Duggan, Kyla. 2013. Dealing with the past: responsibility and personal history. *Philosophical Studies* 164: 141-161.

Fischer, John Martin. 1982. Responsibility and Control. *Journal of Philosophy* 89: 24-40.

- 1987. Responsiveness and Moral Responsibility. In *Responsibility, Character and the Emotions*, ed. Ferdinand Schoeman. Cambridge: Cambridge University Press, 81-106 (reprinted as ch. 3 of Fischer 2006: 63-83).

- 1994. *The Metaphysics of Free Will: An Essay on Control*. Oxford: Blackwell.

- 2006. *My Way: Essays on Moral Responsibility*. New York: Oxford University Press.

- 2012. Semicompatibilism and its rivals. *Journal of Ethics* 16: 117–43.

Fischer, John Martin, Robert Kane, Derk Pereboom & Manuel R Vargas. 2007. *Four Views on Free Will*. Malden MA: Wiley-Blackwell.

Fischer, John Martin & Mark Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.

Frankfurt, Harry G. 1969. Alternate Possibilities and Moral Responsibility. *Journal of Philosophy* 66: 829-839

- 1971. Freedom of the Will and the Concept of a Person. *Journal of Philosophy* 68: 5–20.

Herman, Barbara. 2007. *Moral Literacy*. Cambridge MA: Harvard University Press.

Hieronymi, Pamela. 2004. The Force and Fairness of Blame. *Philosophical Perspectives* 18: 115–48.

- 2006. Controlling Attitudes. *Pacific Philosophical Quarterly* 87: 45–74.

- 2007. Rational Capacity as a Condition on Blame. *Philosophical Books* 48: 109–23.

- 2008. Responsibility for Believing. *Synthese* 161: 357–73.

- 2010. Research Overview, September 2010. [Unpublished ms, available at: http://www.philosophy.ucla.edu/people/faculty/hieronymi/Neither_Nor.html]

- 2011. Reasons for Action. *Proceedings of the Aristotelian Society* 111: 407–27.

- 2014. Reflection and Responsibility. *Philosophy and Public Affairs* 42: 3–41.

Kane, Robert. 2002. Responsibility, Reactive Attitudes and Free Will: Reflections on Wallace's Theory. *Philosophy and Phenomenological Research* 64: 693–98.

Kavka, Gregory S. 1983. The Toxin Puzzle. *Analysis* 43: 33-36.

Korsgaard, Christine M. 2009. *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

- 2010. The Activity of Reason. *Proceedings and Addresses of the American Philosophical Association* 83: 23-43.

Lenman, James. 2006. Compatibilism and Contractualism: The Possibility of Moral Responsibility. *Ethics* 117: 7-31.

Levy, Neil. 2005. The Good, the Bad and the Blameworthy. *Journal of Ethics & Social Philosophy* 1: 2-16.

- 2007. Agents and Mechanisms: Fischer's Way. *Philosophical Quarterly* 57: 123–30.

McGeer, Victoria, and Philip Pettit. 2015. The Hard Problem of Responsibility. In *Oxford Studies in Agency and Responsibility: Volume 3*, ed. David Shoemaker. Oxford University Press, 160–188.

McKenna, Michael S. 2001. Review of J M Fischer & M Ravizza, *Responsibility and Control*. *Journal of Philosophy* 98: 93–100.

Moran, Richard. 2012. Self-Knowledge, 'Transparency', and the Forms of Activity. In *Introspection and Consciousness*, eds. Declan Smithies & Daniel Stoljar. Oxford: Oxford University Press, 211-236.

Otsuka, M. 1998. Incompatibilism and the Avoidability of Blame. *Ethics* 108: 685–701.

Ripstein, Arthur. 2004. Justice and Responsibility. *Canadian Journal of Law and Jurisprudence*, 17: 361-386.

Scanlon, T M. 1998. *What We Owe to Each Other*. Cambridge MA: Belknap Press.

Sher, George. 2005. *In Praise of Blame*. Oxford: Oxford University Press.

- 2006. Out of Control. *Ethics* 116: 285–301.

Smart, J J C. 1961. Free-Will, Praise and Blame. *Mind* 70: 291–306.

Smith, Angela M. 2004. Conflicting Attitudes, Moral Agency, and Conceptions of the Self. *Philosophical Topics* 32: 331–52.

- 2005. Responsibility for Attitudes: Activity and Passivity in Mental Life. *Ethics* 115: 236–71.

- 2007. On Being Responsible and Holding Responsible. *Journal of Ethics* 11: 465–84.

- 2008. Control, Responsibility, and Moral Assessment. *Philosophical Studies* 138: 367–92.

- 2012a. Attributability, Answerability, and Accountability: In Defense of a Unified Account. *Ethics* 122: 575–89.

- 2012b. Moral Blame and Moral Protest. In Coates & Tognazzini (eds) 2012, 27-48.

- 2015. Attitudes, Tracing, and Control. *Journal of Applied Philosophy* 32: 115–32.

Smith, Michael. 2004. Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion. In his *Ethics and the A Priori.* Cambridge: Cambridge University Press, 114–35.

Strawson, Galen. 2008. The Impossibility of Ultimate Moral Responsibility. In his *Real Materialism and Other Essays*, Oxford: Clarendon Press, 319-331.

Strawson, P F. 1962. Freedom and Resentment. *Proceedings of the British Academy* 48: 1-25.

Talbert, Matthew. 2012. Moral Competence, Moral Blame, and Protest. *Journal of Ethics* 16: 89–109.

Vargas, Manuel. 2013. *Building Better Beings: A Theory of Moral Responsibility*. Oxford: Oxford University Press.

Wallace, R Jay. 1994. *Responsibility and the Moral Sentiments*. Cambridge MA: Harvard University Press.

- 2002. Replies. *Philosophy and Phenomenological Research* 94: 707-727.

Watson, Gary. 1975. Free Agency. *Journal of Philosophy* 72: 205–220.

- 2004 [originally 1996]. Two Faces of Responsibility. In his *Agency and Answerability*. New York: Oxford University Press, 260–88.

Williams, Bernard. 1985. *Ethics and the Limits of Philosophy*. London: Fontana.

- 1993. Moral Incapacity. *Proceedings of the Aristotelian Society* 93: 59-70.

Wolf, Susan. 1980. Asymmetrical Freedom. *Journal of Philosophy* 77: 151–66.

- 1987. Sanity and the Metaphysics of Responsibility. In Ferdinand Schoeman (ed), *Responsibility, Character and the Emotions*. Cambridge: Cambridge University Press, 46-62.

- 1990. *Freedom Within Reason*. New York: Oxford University Press.

- 2005. Freedom Within Reason. In James Stacey Taylor (ed), *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*. Cambridge: Cambridge University Press, 258-274.