

Moral Responsibility and Moral Theory – A Package Deal

DRAFT

Elinor Mason

Abstract

In this paper I propose a new way of understanding compatibilist theories of responsibility. I make a distinction between meta views of responsibility (corresponding to meta ethical theories – i.e. theories about how claims about responsibility could be correct) and normative theories of responsibility – theories about when claims about responsibility are correct, and I argue that normative theories of responsibility are not just analogous to normative moral theories, but they are in fact the same theories – I call this the ‘Package View’ – a moral theory is also a theory of responsibility – the two come in a package. I am not proposing new normative theories of responsibility or of morality, rather, I am proposing a way of making sense of, and thus of choosing between, existing theories.

1. Understanding the compatibilist project

Before saying anything else it is worth disambiguating two senses of the phrase ‘free will’. It is sometimes used to refer to something that is not compatible with determinism – agent causation or something like that. This is free will in the strong sense, and when I refer to ‘free will’ in what follows, this is what I am referring to. I start from the position that we do not have free will. Sometimes, however, the phrase

‘free will’ is used by compatibilists to refer to whatever it is that we might have that is compatible with determinism. Hence there is confusion about what compatibilism claims – does it claim that free will is compatible with determinism, or does it claim that responsibility is compatible with determinism? For clarity I will only use the phrase ‘free will’ in the strong sense, and take it that the compatibilist claim is that responsibility is compatible with determinism.

In this paper I set aside the possibility that we have free will. The question of whether we have free will is a question in metaphysics, a question about the nature of causation. My question is not about metaphysics, but about our practices of attributing responsibility to each other. I will assume that we live in a deterministic universe, and that the compatibilist project is to make sense of our practices of attributing responsibility in such a universe.

It is undeniable that our intuitions about moral responsibility are mixed up with intuitions about free will. Similarly, our intuitions about morality involve intuitions about the reality of moral facts. In both cases our pre-philosophical intuitions pull us towards to a form of realism- realism about free will, or realism about moral facts. Philosophers who think that these forms of realism are ruled out on metaphysical grounds must try to give an account of morality, and of moral responsibility that does not appeal to metaphysical facts.¹ Anti-realists must give a new account of our practices.

Just as there are various non-debunking positions an anti-realist about moral facts might take, there are various non-debunking positions an anti-realist about free will might take. One might think that just as there are constructivists and projectivists

¹ In what follows I refer to my starting point as ‘anti-realist’. I do not mean to imply that it is not objectivist. In calling the view anti-realist I mean only that I do not think that any objectivism about moral judgments or judgments about responsibility can be backed up by an independent realm of facts about morality or responsibility.

and so on in meta-ethics, there would be equivalent accounts of the foundations of responsibility in meta-responsibility. To some extent, this is how the landscape looks. There are various meta-responsibility accounts that are more or less equivalent to accounts in meta-ethics (for example, Strawson offers a projectivist account, and Hillary Bok offers a constructivist account). This naturally leads to the thought that, just as there are utilitarians and deontologists and virtue theorists in normative ethics, there would be various accounts of the sorts of thing that people can be responsible for. Again, this is roughly how things are², though the division of labour between meta-theorists and normative theorists is much less sharp in current thinking about moral responsibility than in moral philosophy. In this paper I propose a new way of understanding normative theories of responsibility. I argue that normative theories of responsibility are not just analogous to normative moral theories, but they are in fact the same theories. I should stress that I am not proposing a new normative theory of responsibility, I am proposing a way of making sense of existing compatibilist theories of responsibility.

By a ‘normative theory of responsibility’ I mean a theory about *when* attributions of responsibility are correct. This is opposed to a meta-level theory about how they can be correct at all. This distinction is fairly well recognised and accepted in moral philosophy, but less so in the literature on moral responsibility. The explanation for that is not hard to find: it very easy to slip into realist assumptions when talking about when we should attribute responsibility – it is easy to say things like, ‘we should hold P responsible when he has control of his action’ and so on. Many of the terms that will be relevant to an account of responsibility, such as

² Sometimes, the two accounts are intertwined (as in Wallace, and Bok), sometimes the meta level theory is passed over very quickly in order to concentrate on the normative theory (as in Fischer and Ravizza).

‘capacity’, ‘control’, ‘choice’, ‘ability’, and so on, can be interpreted in metaphysically loaded ways. The compatibilist must beware of what Jay Wallace calls the ‘generalisation strategy’: the sorts of excuse that we accept as defeating responsibility in particular cases appears to generalize to show that if determinism is true then we do not have responsibility at all.³ For example, Bill explains that he was under hypnosis when he killed the cat. A natural thought is that if he was under hypnosis then he could not have done otherwise, and so he is not responsible. The compatibilist has to show that the most basic terms of her theory, as well as her account of exemptions and excuses, are non-question begging. Of course, there is a parallel risk in normative moral theory - it is not acceptable for a claim one shouldn’t lie to be justified by the claim that lying really is wrong.

What do we want from a normative theory of responsibility? First, as I say, the normative theory must be consistent with the anti-realist meta-level theory. The second criterion of a decent theory of when attributions of moral responsibility apply is that the theory *must not be arbitrary*. If we think about human action in terms of a mechanistic universe, we are just part of a huge causal chain. Given that our thoughts and decisions are part of a huge causal chain, which part of the chain are we interested in? Compare a line of dominoes, set up so that when the first one is knocked over, it will knock the second one over, which will knock the next one over, and so on. In one sense it is true that domino 6 is ‘responsible’ for the fall of domino 7, but not any sense that we are interested in. Dominoes are not at all the right sorts of thing for us to care about that undeniable fact that 6 knocked 7 over. Here, the inevitability of 6 having done that entirely swamps any interest we might have in 6 itself. That is not the case with the actions of people – we are very interested in what individual people

³ Wallace, 1994.

do, and the truth of determinism will not affect that interest. This is one of Strawson's main points in his influential piece, 'Freedom and Resentment', and the point is also made by Frankfurt in 'Freedom of the Will and the Concept of a Person'.⁴

How, then, should we pin point more precisely which aspects of the causal chain we are interested in? It cannot be arbitrary that we pin point one aspect of the chain – second order volitions for example, and say that an agent is morally responsible when her act is in accordance with her second order volitions. As critics of Frankfurt have pointed out, we might equally say that an agent is morally responsible when her act is in accordance with her first or fifth order volitions, and there would be nothing to choose between these views. On the other hand, if we say that one aspect of the causal chain is where responsibility *really is* located, then we are assuming something like metaphysical free will. We need a rationale for focussing on some aspect of the causal chain that does not appeal to free will.

Existing compatibilist accounts of moral responsibility are usually extremely cognisant of the first criterion. Philosophers go to great lengths to show that their account of moral responsibility does not contain any question begging terms or concepts, and have offered sophisticated accounts of exemptions and excuses that do not leave the theory vulnerable to the generalisation strategy. However, less attention has been paid to the second criterion. Frequently, philosophers try to show that their account is non-arbitrary by showing that their concept of moral responsibility *just is* our pre-philosophical concept of moral responsibility. For example, Harry Frankfurt argues that moral responsibility does not require alternate possibilities by appealing to thought experiments.⁵ The thought experiments are designed to show that, at least on reflection, we do not believe that alternate possibilities are required for responsibility.

⁴ Strawson, 1962, Frankfurt, 1971.

⁵ Frankfurt, 1969.

Frankfurt's article has generated a huge literature about whether the thought experiments are effective. But we might think that it doesn't really matter. There is no doubt that our reactions to thought experiments such as Frankfurt's are polluted by intuitions about free will, and so the thought experiments may not work to convince us that our existing concept of moral responsibility does not require alternate possibilities. However, if we *start* from the position that we are looking for an account of responsibility that does not depend on free will, we do not need to worry about reactions to thought experiments such as Frankfurt's.⁶

Alternatively, philosophers try to show that the account is non-arbitrary by showing that it matches our intuitions in *most* cases. Although this is no doubt part of what we want from a theory of moral responsibility, it should not be all. A consistently compatibilist theory is bound to depart from our commonsense judgements in some cases, it is bound to be, as Manuel Vargas puts it, revisionist. Compatibilists need a way of defending an account of responsibility *even if it does not* match all our prior intuitions about free will and responsibility. We need an overarching rationale for the normative shape of an account of responsibility. That is my focus in this paper.

2. Wallace's account of Moral Responsibility and Moral Theory

Wallace has several aims in his influential book on moral responsibility. One aim is to integrate a Strawsonian account of the meta-level justification of attributions of responsibility with a robust objectivism about particular attributions. Another is to show that our existing concept of moral responsibility does not require free will.

⁶ Manuel Vargas argues in detail that compatibilist theories will probably have to accept that they cannot capture all our pre-philosophical intuitions in a series of papers. See Vargas 2004 and 2005.

Finally, though he does not say this explicitly, I think that part of his aim is to provide a rationale for a normative theory of moral responsibility, and he does it by arguing that claims about responsibility are dependent on moral norms. I am interested in this last aim because of its obvious similarity to my own project. Although I agree with Wallace that responsibility is not independent of morality, I will argue that he misconceives the connection between morality and responsibility.

There is a simple identification of morality and normative theory that is clearly unsatisfactory. When we ask about whether we should praise or blame someone we are not asking whether we *morally* ought to engage in the act of praising or blaming them (though some consequentialists have made the mistake of thinking that that is the only question)⁷. Ideally, an anti-realist account of responsibility would be able to preserve the commonsense thought that the ‘should’ in ‘should we hold P responsible?’ is a theoretical should rather than a practical one, and that holding P responsible is an attitude with a cognitive element, and not merely an act or non-cognitive attitude.⁸ So I argue for a view according to which we should hold P responsible when it is *correct* to do so, not when it is expedient, or has the best consequences, or is morally right on some other normative view. However, I will show that we can give an account of facts about responsibility without appealing to a realm of independent facts about responsibility. Facts about responsibility are facts

⁷ Philosophers who give a consequentialist account of moral responsibility include JJC Smart (1961), Daniel Dennett (1984). In what follows I ignore the possibility that the consequences of acts of praise and blame should be evaluated independently of attributions of responsibility. I will assume that saying that it is appropriate to praise or blame someone is equivalent to saying that they are responsible. I am not denying that there are disagreements about whether these notions do coincide.

⁸ If no such position can be defended then of course the anti realist will have to give up on the aim of defending the claim that we should hold people responsible when it is correct to do so, and revert to weaker claims.

about morality, more specifically, about the applicability of moral norms. There is nothing metaphysically suspect about these facts.

Wallace argues that the 'should' in 'should we hold P responsible?' is a practical should. Wallace's argument proceeds by focussing on the incompatibilist's suggestion that it would be unfair to hold the agent responsible in the absence of free will. He says:

Philosophers have often suggested that freedom matters to responsibility because it would be unfair to hold people responsible in the absence of freedom of will. I will take this thought much more seriously than it has been taken heretofore, and work it up into an interpretation of what is at stake in the debate between compatibilists and incompatibilists. This normative interpretation has two important consequences. First, we cannot establish what it is to be a morally responsible agent unless we first understand the stance of holding someone responsible—the stance of the moral judge, rather than of the agent who is judged. Second, determining what the conditions of moral responsibility are will require an excursion into normative moral theory; we will need to investigate our principles of fairness, to see what they entail about the conditions under which it would be fair to hold people responsible.⁹

The first part of Wallace's project can be understood as analogous to meta-ethical projects. Wallace's account of the reactive stance is designed to vindicate a broadly Strawsonian position. Wallace agrees with Strawson's view that our attributions of

⁹ Wallace, 1994, p.5.

responsibility are fundamentally motivating¹⁰ but he is uncomfortable with the apparent consequence that there are no standards of correctness for our attributions of responsibility. Obviously there is a parallel here with internalist non-cognitivism in ethics. Just as sophisticated non-cognitivists in ethics attempt to show that there are standards of correctness for judgments of right and wrong, Wallace develops an account of standards of correctness for attributions of responsibility that is consistent with his Strawsonian anti-realism.

I will concentrate on Wallace's account of the standards of correctness.¹¹ As Wallace points out this is a normative question, and we need to know which norms are in play. Wallace argues that we cannot appeal to theoretical norms, as that would amount to assuming that there is a fact of the matter about responsibility independently of our stance of holding people responsible. So, Wallace argues, if the 'should' in 'should we hold P responsible' is not a theoretical should, it must be a practical should. Thus, the argument goes, to determine which substantive norm is relevant here we should look at practical norms – *norms that subserve our practical commitments*, as Wallace puts it.¹² Wallace understands the question, 'should we hold P responsible?' as a moral question, and in particular, the moral question, '*is it fair to hold P responsible?*'. Wallace defends this interpretation of the question by pointing out that the incompatibilist complaint is often that it would be unfair to hold people

¹⁰ See in particular his remarks on p. 66 and p. 78.

¹¹ Wallace sees the two parts of his project as distinct to some extent, but not completely. He says on p.2, "The book may be thought of as having two main parts. One is an account of what it is to hold people morally responsible, in terms of the moral sentiments. The other is an account of the conditions of moral agency, in terms of the rational power to grasp moral reasons and to control one's behavior by the light of them. The leading idea of the book is that these two parts illuminate and reinforce each other, producing together a unified and compelling interpretation of moral responsibility and its conditions." (Wallace, 1996).

¹² P.92. Note that Wallace classes moral norms as practical norms, which is of course, of a piece with his Strawsonian motivational internalism about attributions of responsibility.

responsible if determinism were true. So, we should hold someone responsible only when it would be fair for us to do so. Hence Wallace proceeds by examining our concept of fairness in order to get a grip on the standards of correctness for attributions of responsibility.

This is a very neat argument, but it is flawed. Wallace is wrong to conclude that the only alternative to thinking that there is a realm of independent facts about responsibility is to say that the ‘should’ in ‘should we hold P responsible’ is a practical should.

3. The Package View

Like Wallace I start with the assumption that there are no prior and independent facts about responsibility – in other words, that we do not have strong free will. So how do we decide when P is responsible? There are no metaphysical standards (that’s just our starting point), so we need to look somewhere else. I agree with Wallace that our interest in responsibility is often a moral interest – we are interested in whether people are responsible because we want to know whether we should praise or blame them. However, it is not that we want to know whether we *morally* ought to engage in the act of praising or blaming them.¹³ Unlike Wallace I want to preserve the commonsense thought that the ‘should’ in ‘should we hold P responsible?’ is a theoretical should rather than a practical one.

¹³ In what follows I ignore the possibility that the consequences of acts of praise and blame should be evaluated independently of attributions of responsibility. I will assume that saying that it is appropriate to praise or blame someone is equivalent to saying that they are responsible, and also equivalent to saying that it is appropriate to label their actions right or wrong. Generally I use the expression, ‘apt for moral appraisal’ to cover all of the preceding. I am not denying that there are disagreements about whether these notions do coincide – I intend the coincidence here to be only linguistic and leave open the substantive question til sections 4 and 5.

Our problem is that, in the absence of free will, there seems to be nothing theoretical available to anchor our account of moral responsibility. We might pick any section of the causal chain that leads to an action, and claim that that section is the part that makes an act responsible. But of course that would be arbitrary. In this section I shall show that existing accounts of moral responsibility can be understood as being intimately connected to existing moral theories. This provides a general account of how attributions of moral responsibility are anchored, and I shall show that any compatibilist should accept it, though of course, disagreement may remain about the particular package that we should accept.

One way to get at the question of what it is to be responsible is to ask, what conditions must hold for an act to be apt for appraisal? There are different sorts of appraisal – aesthetic, moral, and so on. The conditions of aptness for these different sorts of appraisal include conditions of *content* as well as conditions that correspond to our notion of responsibility. By ‘content’ I mean the content of the description of the act. Obviously an act can have different descriptions, and which description is relevant is another question. But I think the idea that some contents render an act moral and some do not is fairly clear. For example, only certain sorts of content make an act apt for moral appraisal as opposed to aesthetic appraisal. If an act is an act of ‘adding a small patch of colour to a painting’, it is clearly apt for aesthetic appraisal, but not moral appraisal. The very same act might become apt for moral appraisal if we alter the content of the description – for example, perhaps this act could be described as, ‘destroying the Mona Lisa’ - an act which is certainly apt for moral appraisal, even if only indirectly. Similarly, some moral theories take purely self-regarding acts to be moral, while others do not. We should only morally praise or blame P for her act if it is the right sort of act in terms of content, and likewise, we

should only morally praise or blame P if she is responsible for her act. I will give an account of how we ought to think about responsibility that is analogous to what I take to be a quite uncontroversial view about how we ought to think about these questions of content.

Just as Wallace asks, ‘when should we hold P responsible?’, moral philosophers ask, ‘what sort of content must an act have for it to count as a moral act?’. As with the original question about responsibility, this is a normative question, and so we need to know which norms are in play. Wallace’s argument was that we cannot appeal to theoretical norms, as that would amount to assuming that there is an independent fact of the matter. We certainly should not assume that there is an independent realm of facts making it true that certain content renders an act moral and certain content renders an act aesthetic. So what should we say? Wallace argues that the ‘should’ in ‘should we hold P responsible’ must be a practical should. But even if that seems plausible in the case of responsibility (I shall argue that it is not), it is surely the wrong reading of the question about content. It seems ridiculous to say that we should use moral or other practical grounds to determine whether an act has moral content. It is fairly obvious that we should see the question as a conceptual question – it is not a metaphysical question and it is not a normative question.

Whether or not an act’s content is moral is depends on our concept of morality – it is just like the question of whether something counts as a unicorn, or a table. We can answer the question *without* appealing to a realm of independent facts. There is much that comes from our concept of morality that is shared by all moral theories. An act that has absolutely nothing to do with the suffering of sentient beings – for which no description at all has content concerning sentient beings – does not seem moral whatsoever. However, there is a lot of variation between particular ethical theories,

and our moral concepts start to diverge at some point well before the question of what counts as moral content has been settled. Thus accounts of aptness for moral appraisal will vary with different moral theories. For example, if you are broadly a Kantian, you will tend to think that merely self-regarding actions are apt for moral appraisal, but if you are of some other stripe, you may think that only other-regarding actions are apt for moral appraisal. The relationship between your moral theory and the view you take on what counts as moral content is not one with a clear direction of priority - in fact, the question about what sorts of content are apt for moral appraisal is close to being the same question as the question of what sort of moral theory you have at all - taking a particular view about what counts as moral content isn't simply dictated by the moral theory you have, *it is a large part of having that moral theory*.

The central point of this paper is that we should think about responsibility in the same way. The concept of wrongness has conditions of moral responsibility built into it – it's a package deal. When we say that some piece of behaviour is wrong, we are automatically saying that it is behaviour for which the agent is responsible.¹⁴ Further, different conceptions of wrongness have different conceptions of responsibility built into them. The sorts of things that can be wrong are the very same sorts of things that we can be responsible for. For example, on a Kantian view, intentions are the central focus – they render acts right or wrong, and intentional actions is the only sort of responsible action. The underlying thought is that intentions are *what matter*. The view you take about what counts as responsibility does not just fall out of your moral theory – it is partly constitutive of it.

The traditional landscape in theories of moral responsibility divides up into two sorts of view. On the one hand are 'real self views' – that locate moral

¹⁴ I return to objections to this claim in section 4.

responsibility in the agent's real self. Real self views try to show that we hold (or at least, should hold) an agent morally responsible when her action issues from her 'real self'.¹⁵ The hard question for this sort of view is 'what is the real self?' – what makes an action one that really is the agent's?¹⁶ On the other hand are 'reasons responsiveness views', that assign moral responsibility to agents who have gone through a particular intellectual process in arriving at their decision.¹⁷ Susan Wolf proposes the view that an agent is responsible when she has the ability to act on the basis of right reason (the Good and the True, as Wolf calls it) – the crux of her view is that an agent is responsible when she has the capacity to be governed by reasons. More recently, Hilary Bok and John Fischer and Mark Ravizza, and of course Jay Wallace have argued for reasons responsiveness views, though beyond that structural similarity their views are very different.

The crucial question is, how do we decide between these views about moral responsibility? What are the criteria for deciding between them? There are two well established criteria as I have said. Compatibilists theories must avoid begging the question by using terms that reply on anything like free will. Second, the theory must give an account of moral responsibility that is non-arbitrary. Usually, theorists try to do this by showing that the concept of moral responsibility in their theory just is the same one that we have pre-philosophically. But this is not the only way to meet the non-arbitrariness condition. Compare the way that non-realist theorists in normative

¹⁵ These views are also known as 'structuralist views' (see Fischer and Ravizza, 1998), 'internalist' views (see Fischer and Ravizza, 1998)), and 'Self-disclosure views' (see Gary Watson, 2004).

¹⁶ For a more detailed overview of recent work in moral responsibility see my XXX.

¹⁷ These are often referred to as 'historicist' views, because they use the history of the act to determine whether the agent is morally responsible. Fischer and Ravizza classify them as 'externalist views' (Fischer and Ravizza, 1998). Elsewhere Fischer refers to these as 'reasons responsiveness accounts' ('Recent Work' p. 127).

ethics proceed. There is *some* importance placed on pre-philosophical intuitions about what sorts of things are right and wrong, but it is also understood that a theory that is not realist needs an overarching rationale, and the overarching rationale can lead to quite radical revisionism about morality. So, for example, Neo-Kantians argue about what the fundamental Kantian rationale is – is it that you should not make an exception of yourself? Is it that you should respect humanity wherever you come across it? And they also argue about what is entailed by the fundamental rationale – does it really entail that you must never tell lie, even when confronted with the enquiring murderer?¹⁸ In fact, most contemporary Kantians argue that, contra Kant himself, the Kantian rationale does not require an absolute prohibition on lying, and so are not radically revising pre-philosophical intuitions on this point. But notice that the argument they give for the occasional permissibility of lying is not that it is part of commonsense morality – rather it is that the Kantian rationale permits or even requires lying in some circumstances.

So far we see various parallels between theories of moral responsibility and moral theories. Both are subject to a non-arbitrariness condition. The non-arbitrariness condition could, in theory, be met by showing that each particular directive of the theory exactly matches our commonsense ideas. But in fact our ‘commonsense’ views about morality and moral responsibility are a mess. If we rely on our intuitions about particular cases we will pretty soon a stalemate and not be able to find a way out. So the anti-realist needs another way of showing that the theory of morality, or of moral responsibility, is not arbitrary. The non-arbitrariness condition is met by the overarching rationale.

¹⁸ See for example Korsgaard’s discussion in Korsgaard (1996), and Barbara Herman 1993.

This is taken for granted, if rarely made explicit in normative moral theory. However, it has not been applied to compatibilist theories of moral responsibility. The reason that compatibilists are so wedded to our existing concept of moral responsibility is that they don't have anything else to go on – they don't have an overarching rationale for an account of moral responsibility. That is what I aim to provide in this paper. This is where we can move beyond saying that there are parallels between theories of moral responsibility and normative moral theories: the overarching rationale for a particular view about moral responsibility *is just the same as it is for the rest of the corresponding moral theory*. The right sort of theory to have is a package of claims about responsibility and morality, united by one overarching rationale.

I said that theories of moral responsibility can broadly be divided into real self views and reasons responsiveness views. We can make a similar distinction between different sorts of morality: character focussed moralities, and intention focused moralities. Aristotle's theory, and related contemporary virtues theories are the best examples of character focussed theories. Virtue theories are united by the thought that what really matters is not primarily the acts that you do, but the character that you have. Ethically, it is important to be a good person, in a deep and stable way. Obviously, this is related to the real self view of responsibility, according to which you are responsible when you are acting from your deep self. The two sorts of view are both coming from the same place: the thought that what really matters is deep self. In normative ethics, the claim that the real self is morally important is not taken to be mysterious, and nor should it be when the same claim is made about moral responsibility. Real self views of responsibility should claim that real self matters for responsibility because it is *morally* important. Morally important in exactly the same

way that it is for normative theory – that basic thought provides the foundation for real self views of morality and real self views of moral responsibility.

Obviously, intention focused moralities share an overarching rationale with reasons responsiveness views about responsibility. Kant's theory and contemporary deontologies are the best examples of an account of intention focussed morality – I'll come back to consequentialism, which, I shall argue, is also in this category. When theorists like Fischer and Ravizza try to come up with an account of moral responsibility, they are partly trying to justify a certain kind of morality. Why come up with a theory of moral responsibility according to which we act responsibly when we act intentionally? Because we already have the idea that intentions are what matter. The overarching rationale of a particular package provides a justification for normative judgments about moral responsibility and morality. The overarching principle is what satisfies the non-arbitrariness condition on an account of moral responsibility – and it takes the place of a metaphysically suspect concept of responsibility.¹⁹

Interestingly, in recent work on both normative theory and moral responsibility there are signs of a convergence of the two sorts of view. Each is anxious to capture what seems right and interesting about the other. So, in normative ethics, Kantians and consequentialists work hard to accommodate the thought that character matters,²⁰ and meanwhile virtue ethicists work hard to show that their view

¹⁹ Particularists, of course, would argue that there are no general principles like my overarching principles. But notice that particularists tend to be realists of some sort or other. And that is no coincidence – realism meets the non-arbitrariness condition in a different way. So I deny that particularists have a challenge here that is independent of a challenge to anti-realism.

²⁰ For example (there are too many to list here) both Herman (1983) and Baron (1991) give accounts of how a Kantian can deal with issues of motivation. Julia Driver (2001) gives a consequentialist account of the virtues.

can be action guiding and deliver verdicts of rightness and wrongness.²¹ In discussions of responsibility, real self views try to avoid the criticism that they capture a merely causal relationship between the agent and her act by incorporating more of the reasons responsiveness view.²² On the other hand, reasons responsiveness views seem problematic because they don't say enough about what makes the act the agent's own—so they can be improved by adding more about what makes an act an agent's own, which of course brings them closer to real self views.²³ The parallel convergences are not surprising. What makes a pure intention view seem worrying in normative ethics is, of course, the very same thing that makes the view worrying as an account of responsibility: we can't help thinking that intentions don't always reflect character, and character seems to matter too. Likewise, what makes a deep self view worrying in one field makes it worrying in the other too – character matters, but intention matters too. The convergence of intention and deep self views is grist to my mill.²⁴

The Package View says nothing about whether our attributions of rightness and wrongness or of responsibility are truth apt in any straightforward way. The meta-ethical project still remains. What my account shows is that our theory of responsibility and our moral theory stand or fall together – there is no special problem of moral responsibility. Similarly, the Package View does not show that our normative moral theories (and hence our theories of moral responsibility) can be free of pre-philosophical intuitions. The overarching rationale for the package is inevitably

²¹ See particularly Hursthouse, 1999.

²² Watson's remarks about practical identity, (2004b) and Bratman's interpretation of identification might be interpreted in this light

²³ Fischer and Ravizza's explanation of what it is for an agent to own her action, and their talk of 'taking responsibility', for example, are reminiscent of Frankfurt.

²⁴ Watson (2004b) has a very different account of the relationship between real self views and intention views, I discuss his account below.

something that we have invented by reflecting on our intuitions. And no doubt some of the original intuitions were polluted by realism of various sorts – realism about moral facts, or realism about responsibility. Both of these have to be rooted out in the development of the overarching rationale, in order to provide a non-arbitrary basis for an account of right and wrong.

In sum, the point of connecting theories of moral responsibility to moral theories is to give them a non-metaphysically suspect rationale. The rationale is the broad ethical outlook, and that outlook grounds attributions of rightness or wrongness. In moral theory, it is widely accepted that some sort of rationale is needed. In theories of moral responsibility, this point has been obscured, but I have argued that a compatibilist normative account of moral responsibility (i.e. an account of *when* we are responsible, as opposed to an account of how we can be responsible at all) is equally in need of some sort of grounding. Why think that the rationale is the same as that for moral theory? Because when we discard the metaphysical aspects of our concept of moral responsibility, what is left is that moral responsibility is a condition of moral appraisal, and so the place to look for a rationale for the conditions of moral appraisal is obviously, morality. When we do look there, what we find is that existing theories of moral responsibility correspond very closely to existing moral theories. And this is no coincidence, we are after all, looking for *moral* responsibility. Theories of morality and moral responsibility have the very same foundation in a thought about what part of our behaviour really matters.

I should stress that existing theories will not fall perfectly into these categorizations. It may be that some theories have not been properly understood even by their authors, and that seeing them in the light of the correspondence between moral theories and theories of moral responsibility will help to tidy up those theories.

Alternatively, it may be that some theories are hybrid. One of the advantages of seeing the identity of normative theories of morality and responsibility is that we can now criticize these hybrid theories. I return to the uses of the Package View in section 6.

4. Isn't there a general sense of responsibility that does not imply rightness/wrongness?

It is a central tenet of the Package View that wrongness and responsibility cannot be separated. If an act is wrong, we can deduce that the agent is blameworthy. If the agent is blameworthy for doing something, we can deduce that her act was wrong. Of course, this issue gets muddled by different terminologies. The term 'blame' is notoriously difficult, appearing in contexts where it refers to attitudes only, others where it refers to acts only,²⁵ and sometimes people use the term 'wrong' to refer to acts that have nothing to do with agency.²⁶ Throughout I have been using the terms 'blameworthiness' and 'praiseworthiness' to refer to the appropriateness of taking up attitudes of praise and blame, and have taken this to be equivalent to seeing someone as responsible. I have said nothing about which if any acts of praise or blame are appropriate, or what standard those acts should be judged by. I have been using the term 'wrongness' to refer to acts that ought not to be done, and I have assumed that we tend to use the word in such a way that it does imply responsibility (and hence blameworthiness or praiseworthiness).

²⁵ And it is even used in contexts where, (in Parfit's terminology, 'blameless wrongdoing'), it refers to the status of an act that is suboptimal on one way of adding up the value of consequences and not suboptimal according to a different way of adding up the value of consequences.

²⁶ These uses of the terms are broader than my use. If the broader uses are accepted, we will still want to find a way to talk about the narrower sense of blame which is conceptually connected to wrongdoing, and the narrower sense of wrongdoing which is connected to blameworthiness. Darwall makes the same point (Darwall 2006 p.95).

Verbal disputes aside, there is a substantive issue here. On my view, moral responsibility is so closely wedded to morality that we are only responsible for acts that are moral acts. Responsibility is a moral notion, and there is no general sense of responsibility. The other side of the coin is that any account of wrongness must have an account of responsibility built in, and so any attribution of wrongness implies an attribution of responsibility. I will return to that issue in the next section.

An objector is bound to point out that the Package View leaves no room for actions that are neither right nor wrong nor praiseworthy nor blameworthy but that the agent is responsible for. Or, to put it another way, the objection is that responsibility cannot be explicable in terms of rightness and wrongness because we have a more general conception of responsibility, according to which you might be responsible for entirely non-moral actions, like what side of bed you got out of. On the standard picture, moral responsibility piggybacks on general responsibility, not the other way round.

In one sense the objection can be given short shrift – insisting that there is a prior sense of responsibility is just begging the question against the compatibilist. It is often the case that objections to compatibilist accounts of responsibility amount to a complaint that the account does not capture our intuitions about free will. Naturally a compatibilist account of responsibility is not going to capture every aspect of our commonsense concept of responsibility – it must be to some extent, revisionist. Nonetheless, there might be something more serious in this objection – it might be an objection to the particular way in which the Package View is revisionist. The objection might be that existing compatibilist accounts can make sense of cases where there is responsibility but moral blame is not appropriate. It does seem that there are actions that we take people to be responsible for, but do not consider moral. It seems

that there is a more general sense of responsibility. Gary Watson gives the example of someone who stays up too late before an important work event the next day, and thus underperforms.²⁷ In this situation we take the agent to be responsible for his own action – he was not coerced, or acting out of mental illness. However, we do not think it appropriate to blame him for his action.

Watson's own solution here depends on his distinction between attributability – attributing an act to an agent, and accountability, holding an agent accountable. According to Watson, accountability is the conception that runs into trouble with determinism, because our ideas about accountability are mixed up with our ideas about avoidability, and it is our ideas about avoidability that appear to be incompatible with determinism. According to Watson, we will only hold the agent accountable for *some* of the things that they really did. On Watson's view, attributability is the deep notion of responsibility, and accountability is the social notion. Thus the agent who stays up too late the night before an important performance is responsible in the sense of attributability, but she is not accountable to us.

Watson hints that his distinction between attributability and accountability corresponds to the distinction between real self views and reasons responsiveness views.²⁸ However, though Watson has certainly identified two aspects of our concept

²⁷ Watson 2004b.

²⁸ Watson's article grew out of a reply to Susan Wolf, who accuses real self views of being shallow – of not capturing the deep sense of responsibility. Watson argues that the real self view is not shallow – that in fact it captures what is crucial in responsibility: “The point of speaking of the “real self” is not metaphysical, to penetrate to one's ontological center; what is in question is an individual's fundamental evaluative orientation. Because aretaic appraisals implicate one's practical identity, they have an ethical depth in an obvious sense... To adopt an end, to commit oneself to a conception of value in this way, is a way of taking responsibility... Hence one notion of responsibility – responsibility as attributability – belongs to the very notion of practical identity.” (Watson 2004b p. 271) This is a good

of responsibility, they do not correspond to the distinction between real self views and reasons responsiveness views. On both real self views and reasons responsiveness views it is possible to identify acts that an agent is responsible for that we do not hold her accountable for: an agent whose act is really hers (in whatever the relevant sense is) might not be accountable to us for that act, and likewise, an agent who really reasoned to her act (again, in whatever the relevant sense is) might not be accountable to us for that act. So Watson's distinction between attributability and accountability cuts across the distinction between real self views and reasons responsiveness views. In fact there are three senses of responsibility here – the general or neutral sense of responsibility, moral responsibility, and then accountability, which is concerned with sanctions.²⁹ My theory rules out the general sense, but not the other two. I admit that there is a persistent intuition that there is a general sense of responsibility – and that it seems that Watson's late night reveller is responsible in just this sense. But the Package View leaves no room for that claim.

I think that this objection can be answered. First, in saying that there is only a moral sense for responsibility I am not saying that all acts that are responsible ought to be rewarded/punished. The claim is simply that such acts are apt for the attitude of praise or blame – punishment is a different issue. Second, we do not have perfectly good existing accounts of the general sense of responsibility. I have argued that there

example of the convergence between real self views and intellectual process views that I mentioned earlier.

²⁹ Stephen Darwall and Jay Wallace both give accounts of moral responsibility that tie it very closely to morality. Both claim to be developing accounts of accountability rather than attributability. However, accountability is just about when sanctions are appropriate, and that is a purely moral question – so it is not surprising at all that accountability can be explained in moral terms. The more ambitious aim would be to explain attributability in moral terms – that, of course, is my aim. I suspect that despite what they say it is both Wallace's and Darwall's too, and so they, like me, will end up open to the objection that there is a more general sense of responsibility that cannot be explained on their views.

is something missing from existing compatibilist theories of responsibility. They attempt to pin point something in the causal chain and tell us that that is where responsibility is located. However, they attempt to do this in a piecemeal way, just by trying to get a good match with our pre-philosophical intuitions about responsibility. I have provided a theoretical rationale for compatibilist theories of morality. In linking moral responsibility to moral theory we get a rationale, but of course we lose the independence of responsibility, and we have to deny that there is a sense of attributability that is independent of moral responsibility.

I do not think that this is damaging. For a start, it is worth emphasising that moral responsibility is not the only sort of responsibility - moral theory is not the only sort of normative theory. We might also have accounts of aesthetic, epistemic, and even prudential responsibility. Different accounts of aesthetics will give different accounts of aesthetic responsibility, and so on. If we accept a normative theory of prudence, then we have a corresponding notion of prudential responsibility. It is very likely that, one person's normative theory of prudence would be similar to their normative ethical theory in its structural aspects, and differ only in content. (It is unlikely, but not impossible of course, that someone would hold a consequentialist moral theory and a deontological prudential theory). So prudential responsibility would be just like moral responsibility, and differ only in the content of the responsible act. Thus we have an answer to Watson's case – the person who stays up late before an important performance is not morally responsible, but they are prudentially responsible. Because prudential responsibility and moral responsibility are likely to be structurally similar, it is easy to see how we could get confused between the two.

The objector is bound to press on, and argue that there are cases where existing compatibilist theories can make attributions of responsibility independently of any moral, aesthetic, epistemic or prudential theories, and my understanding of existing theories renders them incapable of doing that. But there is a tempting ersatz use of ‘responsibility’, and that this is what explains apparent cases of responsibility floating free of normative theory. Let me start with a parallel case. Imagine that Johnnie is a plastic surgeon specialising in Botox. One day he accidentally jabs the needle in too far, unintentionally performing a partial lobotomy. Unbeknownst to him, his client is a serial killer, and the partial lobotomy has the effect of rendering the client sweet and good natured – there will be no more killing. Has Johnnie acted rightly? Strictly, his act is not apt for moral appraisal, and so of course he does not act rightly. But we would understand what people were saying when they said things like, it’s a good thing that Johnnie did what he did; Johnnie did something right for once, Johnnie did the right thing there; we really should praise him for what he did, and so on. Johnnie’s act fails to meet some of the conditions for aptness for moral appraisal (the conditions that correspond most closely to our concept of responsibility), but it meets others (the outcome is good for sentient beings). Terms like ‘rightness’ and ‘praise’ here are obviously inappropriate, strictly speaking, but it is easy to see how confusion arises. My case is implausible of course, but confusions between appropriate and ersatz uses of the term ‘right’ are rampant in the literature on probable and actual consequences consequentialism. So if there is a case in which some of the conditions for responsibility are met, but the content of the act does not relate to any normative theory, then it is not surprising that people will tend to an inappropriate use of the terms. In short, I am biting the bullet here. There is no

appropriate attribution of responsibility in the absence of a more general normative theory of some sort. This is revisionist, but it is not immoderately revisionist.

5. Isn't there a sense of rightness/wrongness that does not imply responsibility?

From the other direction, an objector might insist that I should not have tied wrongness so closely to responsibility. On one construal of consequentialism, the theory says that an action is right if it maximises the good. This is often known as 'objective' or 'actual consequence' consequentialism – rightness depends on actual consequences, and right and wrong actions have no necessary connection to agency – they may be completely inaccessible to the agent. I'll deal with consequentialism in more detail below. In the meantime, there seem to be numerous other cases that we could construct where our ordinary language uses the terms rightness and wrongness in such a way that they diverge from responsibility: two people both push someone over and break that person's leg but one person is hypnotised, the other just malign. So, it is tempting to say, they both did something wrong but only one is responsible. The basic problem is that we are conflating badness and wrongness, and there is no reason to think that this is anything other than a verbal dispute. We can all agree that within the category of bad events we can distinguish bad things that happen that an agent is responsible for, bad things that happen that are caused by an agent but for which she is not responsible, and bad things that happen that have nothing to do with an agent. Why not keep these ideas distinct? Why insist that both of the first two must be called wrong acts? Objective consequentialism is the only theory that seriously makes the claim that the second sort of case really does describe wrong acts in some important sense, and so I will return to that now.

6. What is the point of the Package View?

I have argued in this paper that the overarching rationale for a theory of responsibility is not just parallel to that for moral theory - it is the same. Inevitably none of our concepts from moral theory will perfectly capture all our intuitions about responsibility - inevitably some of our concepts will depend on false metaphysical views, and others will imply false or incoherent views. The Package View implies that we must use a process of reflective equilibrium to come to an account of moral responsibility that matches as well as possible our ideas about morality, and an account of morality that matches as well as possible our ideas about moral responsibility. We can learn about both morality and responsibility by being attentive to the interdependence. In closing I will briefly discuss a few of the ways in which the Package View changes the way we should argue about responsibility and morality.

First, I have already mentioned objective consequentialism. Many philosophers think that there is something very wrong with a picture according to which an agent can act rightly entirely accidentally. Philosophers have struggled to put this argument in clear terms. My account of the connection between moral theory and moral responsibility makes sense of the unsatisfactoriness of objective consequentialism.³⁰ What the objector to actual consequence consequentialism is trying to express is the thought that a moral theory should embody a decent account of moral responsibility. Actual consequence consequentialism does not – it is wildly implausible to say that you are morally responsible for every causal consequence of your action. Such theories do not provide the right sort of connection between the agent and the things he is responsible for - a merely causal connection is not enough

³⁰ For a more detailed account of this see my XXX.

to capture the ‘deep sense of responsibility’ that even compatibilists are entitled to look for.

Of course, the objective consequentialist will reply that she is not trying to give an account of moral responsibility. My response is that she should have been, or perhaps even that she must have been without realizing it. The best versions of consequentialism are subjective probable consequence consequentialisms, that acknowledge that there must be a robust and interesting connection between the agent and the actions that we can correctly label right and wrong. Such theories belong in the category of intention focused theories. So there is room for consequentialism after all, but perhaps not where it might have been expected. It turns out that the big divide in ethical theories is not between consequentialism and deontology, but between consequentialism and deontology on the one hand and virtue ethics on the other. The more general lesson here is that moral theories must be concerned with agents. That is not to say that moral theories cannot talk about value in things other than agents – in sunsets and avalanches and so on. However, a moral theory cannot talk *only* about non-agents – and that is the mistake that objective consequentialism embodies. A crucial part of a moral theory is about the rightness of agents’ behaviour, and that is where responsibility comes in – the only sorts of actions that can be right are the sorts of actions that agents can be responsible for.

Consequentialist theories are the class of theories most in need of tidying up through the Package View. But this is not the only way in which the Package View should change our thinking about normative moral theories. Take the discussions of ‘moral luck’ in the literature. Once we have the Package View in mind, it is clear that Nagel’s and Williams’ discussions of moral luck are discussions of the way in which the intention focussed package can be criticised from the point of view of the

character focussed package. Likewise Adam's discussion of involuntary sins in his paper of the same name. Seeing these arguments in terms of the Package View clarifies what is going on – there is no puzzle of moral luck – there are just different accounts of responsibility/morality – different packages. Or take Nomi Arpaly's claim that Huck Finn style examples – examples where someone does something good while believing it to be the wrong action – show that autonomy is less important to morality than has been thought. With the Package View firmly in mind we can see that this claim is not an independent challenge to intention focussed packages, rather it is just an expression of support for a character focussed package.

Equally, theories of responsibility benefit from being conceptualised as part of a package. Rather than designing endless permutations of examples and counterexamples to test intuitions about responsibility, theorists about responsibility should be looking at the bigger moral picture. Why does intention matter? Why does real self matter? With the bigger picture as a justifying tool certain apparent counterexamples will not matter at all, and some will matter less. The Package View provides a whole new set of tools

I mentioned the possibility of 'hybrid' theories earlier. It might seem that someone could propose a theory that respects the general constraint – that rightness and wrongness are tied to responsibility, without using the same theory for rightness as for responsibility. An intention focussed morality such as subjective consequentialism might be paired with a real self account of responsibility, or a character focussed morality might be paired with reasons responsiveness account of responsibility. It is immediately clear that there is something a bit odd about such theories. Imagine a virtue ethical view according to which the right (or virtuous) act is one which expresses the agent's settled character, but the acts for which she is

responsible are the ones that has reasoned to. Those acts might not be the same acts! So in order to respect the general constraint that rightness and responsibility have to converge, the class of acts that are right/responsible must be only the intersection of the two criteria – i.e. the acts which both express the agents settled character and are reasoned to in the right way. But there is something quite odd about that – it begs the question, if rightness is all about settled character, why do only some of the acts that express settled character count as right? My account of the bigger picture here explains why hybrid views are unsatisfactory: they contain conflicting accounts of what ultimately matters.

Finally, the Package View suggests new avenues for thinking about responsibility – as well as moral responsibility there are other sort of responsibility – other normative packages. I mentioned epistemic responsibility and aesthetic responsibility – there may be others. The Package View gives us a new way to start thinking about these questions.

Bibliography

Adams, Robert (1985) 'Involuntary Sins' *Philosophical Review* 94 (1):3-31.

Aristotle 1985: *Nicomachean Ethics*, translated by Terence Irwin. Indianapolis: Hackett Publishing.

Arpaly, N. (2000) "On Acting Rationally Against One's Better Judgment," *Ethics* 110: 488-513.

Marcia Baron 1991: 'Impartiality and Friendship', *Ethics* 101, pp. 836-57

Bok, Hilary 1998: *Freedom and Responsibility*. Princeton University Press.

Bratman, Michael 1999: *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge: Cambridge University Press.

-----1999b: 'Identification, Decision and Treating as a Reason', reprinted in Bratman 1999.

----- 2006: *Structures of Agency*. Oxford: Oxford University Press.

Darwall, Stephen 2006: *The Second Person Standpoint*. Cambridge, Mass. Harvard University Press.

Dennet, Daniel C. 1984: *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, Mass.: MIT Press, 1984.

Driver, Julia 2001: *Uneasy Virtue*. Cambridge: Cambridge University Press.

Fischer, John Martin 1999: 'Recent Work on Moral Responsibility.' *Ethics* 110: 93-139.

----- 1994: *The Metaphysics of Free Will*. Oxford: Blackwell Publishers.

----- ed., 1986. *Moral Responsibility*. Ithaca: Cornell University Press.

Fischer, John Martin and Ravizza, Mark 1998: *Responsibility and Control: An Essay on Moral Responsibility*. Cambridge: Cambridge University Press.

Frankena, William 1950: 'Obligation and Ability', in Max Black, (ed.) *Philosophical Analysis: A Collection of Essays*. London: Prentice-Hall.

Frankfurt, Harry 1999: *Necessity, Volition, and Love*. Cambridge: Cambridge University Press.

----- 1988: *The Importance of What We Care About*. Cambridge: Cambridge University Press.

----- 1987: 'Identification and Wholeheartedness.' Reprinted in Frankfurt, 1999.

----- 1971. 'Freedom of the Will and the Concept of a Person.' *Journal of Philosophy* 68: 5-20. Reprinted in Fischer, ed., 1986; Frankfurt, 1987; and Watson, ed., 1982.

----- 1969. "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 46, pp. 829-839. Reprinted in Fischer 1986 and Frankfurt 1987.

Herman, Barbara 1993: *The Practice of Moral Judgment*. Cambridge, MA: Harvard

University Press.

-----1983: "Integrity and Impartiality," *The Monist* 66 (1983), pp. 233-249.

Honderich, Ted 1988: *A Theory of Determinism*. Oxford: Oxford University Press.

Hume, David 1978: *An Enquiry Concerning Human Understanding*. Ed., P.H. Niditch. Oxford: Clarendon Press.

----- 1978: *A Treatise of Human Nature*. Ed., P.H. Niditch. Oxford: Clarendon Press.

Hursthouse, Rosalind 1999: *On Virtue Ethics*, Oxford: Oxford University Press.

Kane, Robert (ed.) 2003 *Free Will*, Blackwell.

Kant Immanuel 1993: *Critique of Practical Reason*, tr. by Lewis White Beck. Upper Saddle River, NJ: Prentice-Hall Inc.

-----1989: *Groundwork of the Metaphysics of Morals*, tr by H.J. Paton. Routledge.

Korsgaard, C. 1996: *Creating the Kingdom of Ends*, Cambridge: Cambridge University Press.

Mason Elinor 2005: 'Recent work on Moral Responsibility' *Philosophical Books* 46,

pp. 343-353.

Nagel, T. 1979 'Moral Luck' in *Mortal Questions*, New York: Cambridge University Press.

Norcross, Alastair 1997: 'Good and Bad Actions' *The Philosophical Review*, Vol 106, (1) pp. 1-34.

Pereboom Derk 2001: *Living Without Free Will*, Cambridge: Cambridge University Press.

Smart J.J.C. 1961: 'Freewill, Praise and Blame'. *Mind* 70, pp. 291-306.

Smilansky Saul 2000: *Free Will and Illusion*. Oxford: Clarendon Press.

Stocker, Michael 1971: "'Ought' and 'Can'", *Australasian Journal of Philosophy* 49, pp. 303-316.

Strawson P.F. 1962: 'Freedom and Resentment.' *Proceedings of the British Academy* 48, pp.187-211. Reprinted in Watson, ed. 1982.

Vargas Manuel 2004: 'Responsibility and the Aims of Theory: Strawson and Revisionism', *Pacific Philosophical Quarterly* 85, pp.218-241.

----- 2005: 'The Revisionists Guide to Responsibility', *Philosophical Studies* 125, 399-429.

Velleman, J. David 1989: *Practical Reflection*, Princeton: Princeton University Press.

-----2000: *The Possibility of Practical Reason*. Oxford: Oxford University Press.

Wallace R.J. 1994: *Responsibility and the Moral Sentiments*', Harvard University Press.

Watson, Gary 2004: *Agency and Answerability* (Oxford University Press, 2004).

-----2004b: 'Two Faces of Responsibility', in Watson 2004.

-----2003 (ed.) *Free Will*, 2nd edition, Oxford: Oxford University Press.

-----2001: 'Reason and Responsibility.' *Ethics* 111: 374-94.

----- 1987. 'Responsibility and the Limits of Evil: Variations on a Strawsonian Theme'. Reprinted in Fischer and Ravizza, eds., 1993.

----- (ed.) 1982 *Free Will*, Oxford: Oxford University Press.

Wolf, Susan, (1990), *Freedom within Reason* (New York: Oxford University Press).

Williams, B. (1981) 'Moral Luck' in *Moral Luck*, Cambridge: Cambridge University

Press