

# Consequentialism and Moral Responsibility

*Draft of September 2015*

Elinor Mason

For Christian Seidel (ed.) *Consequentialism: new directions, new problems?* OUP, forthcoming.

There are two different ways of thinking about the relationship between consequentialism and moral responsibility. First, we might think that consequentialism can give us an account of responsibility. I discuss this possibility briefly, and then set it aside. The other way of thinking about the relationship is the focus of this paper. The question that concerns me, is, to what extent is a normative theory, consequentialism in particular, constrained by requirements that stem from concerns about responsibility?

## 1. Consequentialist Accounts of Moral Responsibility

J.J.C. Smart suggests that we can extend consequentialist reasoning about morality to reasoning about responsibility. One of the attractions of consequentialism is that it provides such a straightforward and attractive account of justification for our moral practices. Why do we pay our taxes, treat each other with respect, look after each other and so on? Because doing so has good consequences. However, this sort of justification, though very appealing when considering moral practice, becomes extremely counterintuitive in other sorts of case. For example, it seems obvious that justification for beliefs cannot be consequentialist. Beliefs must be justified in some way that relates to their truth, though of course there is disagreement about exactly what makes a belief justified. Similarly, so a familiar line of thought goes, whether or not someone is responsible for an act, or for anything else, cannot be determined by looking at the consequences of holding them responsible. The claim that

responsibility can be understood in a consequentialist way seems like a category mistake.<sup>1</sup>

Smart's view might be correct that, insofar as praising and blaming are *actions*, consequentialists should take the value of the consequences of performing those acts as the relevant factor in deciding whether or not to perform them. But in the more important sense of praise and blame (or so it seems), praise and blame are not acts but *attitudes*, and thus, are like beliefs, in that giving consequentialist reasons for them seems mistaken. I should believe that there is a tree outside my window because that is what my evidence suggests – there is a fact of the matter, and I am rational to believe that there is a tree. Likewise, one might think, there is a fact of the matter about whether someone is responsible, and the consequences of pointing out that fact are not relevant to its truth.

Things are slightly more complex than this. One might not be a realist or even an objectivist about responsibility: one might think that 'facts' about responsibility are constructed, or projected. Nonetheless, there seems to be a fundamental error in giving consequentialist reasons for taking someone to be responsible. I shall not pursue this further, but shall simply agree that consequentialist accounts of responsibility are on the wrong track. We need an account of responsibility that makes better (although perhaps not perfect) sense of our pre-theoretical intuitions about the sort of thing responsibility is.

## 2. The Responsibility Constraint

The question I am concerned with here looks at the issue from the other direction. Consequentialism cannot be used to give an account of responsibility, but maybe the two are not entirely independent. What constraints come from our concern with moral responsibility? One might think that consequentialism has nothing to do with responsibility. Consequentialist theories have, it is commonly said, two parts, a theory of the good and a theory of the right. It is plausible that theories of the good have nothing to do with responsibility. What's good is good independently of who or what is responsible for it. Sunny days are good. But an account of rightness seems to have

---

<sup>1</sup> See Smart 1963 for the consequentialist account of moral responsibility, and too many commentators to list on the problem with this view. A more moderate consequentialist account of responsibility has recently been defended by Manuel Vargas (2013), but his view is not that each act we are responsible for is determined by consequentialist considerations, it is rather that the practice of responsibility itself is justified by broadly consequentialist considerations.

some necessary connection to notions of moral responsibility, though it is not clear precisely what the connection is.

An account of the right, as opposed to an account of the good, is more than mere appraisal. It does not just describe value. Rather, it tells us what to do, and it assesses our behavior, *as agents*. This is not very precise, but it is enough to see that there is a constraint on normative theories that does not apply to theories of the good. To see that the constraint clearly applies to consequentialism, consider a case where someone is drowning in a pond a few miles away from me, and I see them through my telescope. My theory of value tells me that it would be bad for the person to drown, better for them to live.<sup>2</sup> But what does my normative theory tell me? On one simple view of consequentialism, it tells me to do what will have the best consequences. So I should teletransport to the pond and save the person. Of course, no-one thinks that consequentialism says that – we all agree that the class of things that I can be asked to do is limited by what is possible for me. A version of this thought has been called the ‘ought implies can’ constraint’ – an ‘ought’ only applies if the agent can do what is being asked. Unfortunately, the sense of ‘can’ in ‘ought implies can’ is not at all clear. Consider another, familiar, example: the only way to save someone is to beat Karpov at chess.<sup>3</sup> I can beat Karpov, in the sense that I can physically make the moves, but I cannot in a more important sense – I don’t know how.

So ‘ought implies can’ is not a useful way to capture this thought, but there is surely a cogent thought here. A normative theory is limited in what it can tell people to do. It is not limited by what is possible for an agent - the notion of possibility is too weak, as the Karpov example illustrates. Another discussion from the literature on consequentialism focuses on a closely related point. Both J.J.C. Smart and Shelly Kagan discuss the ‘epistemic objection’, a worry about the unknowability of future consequences. Both try to answer this by saying that we can focus on the near consequences and ignore the distant ones.<sup>4</sup> So again, what is possible for us is not something that seems apt as the focus of a normative theory: we can and do causally produce states of affairs in centuries ahead of us, but there is something very odd about saying that rightness is defined by these distant consequences.

---

<sup>2</sup> We can say for the sake of argument it is a good person.

<sup>3</sup> From Howard-Snyder 1997, discussed in Mason 2003. See also Miller 2003, Wiland 2005.

<sup>4</sup> See e.g. Smart 1973, Kagan 1998. See also Norcross 1990 and Jackson 1991.

There is clearly a conceptual constraint on what normative theories can look like. Right action must be something that is not just possible for us, but something we can reasonably be deemed responsible for. This is clearly the driving force behind various objections to consequentialism. It is not clear yet what the responsibility constraint involves, and that is what I shall be discussing in this paper, but it is worth a rough and non-controversial formulation:

**The Responsibility Constraint:** A normative theory must give an account of right action such that an agent could reasonably be deemed responsible for acting rightly or wrongly.

This obviously rules out accounts of consequentialism that instruct the agent to do things that are not possible for the agent, such as teletransport. It may also rule out accounts of consequentialism that allow all causal consequences to be relevant to rightness – rightness cannot depend on consequences I do not know about or cannot control, because I cannot be responsible for things I do not know about or cannot control. Many consequentialists have embraced this conclusion, arguing that instead of objective consequentialism<sup>5</sup>, we should opt for prospectivism<sup>6</sup>, or even subjectivism.<sup>7</sup> Alastair Norcross for example, takes it as obvious that there is a responsibility constraint on normative theories: “there may be a temptation to abandon consequentialist methods of reasoning on the grounds that what one is able to bring about, or able to be certain to bring about, is somehow dwarfed by the scale of what else is beyond one’s control or beyond one’s epistemic reach. In each case the cure is to focus on what is within one’s control.”<sup>8</sup> Frank Jackson uses the responsibility constraint to argue for a decision theoretic form of consequentialism, or

---

<sup>5</sup> By ‘objective consequentialism’ I mean the view that right action is defined in terms of the value of the consequences, rather than in terms of the agent’s predictions. The label comes from the fact that right action is defined in terms of what is objectively the case, as opposed to what is the case from the agent’s subjective perspective, which gives us ‘subjective consequentialism’. The view might equally be called ‘actual-consequence consequentialism’ (as opposed to expected consequence consequentialism) or, as Julia Driver (2011) aptly labelled it, the distinction between ‘evaluational externalism’, which bases rightness on factors that are external to agency, and ‘evaluational internalism’, which bases rightness on factors internal to agency.

<sup>6</sup> Defenders of prospectivism include Bentham 1789, Mill 1863, Smart 1973, Gruzalski 1981, Jackson 1991, Oddie and Menzies 1992, Timmons 2002, Howard-Snyder 1997, Zimmerman 2008.

<sup>7</sup> See Fred Feldman 2012.

<sup>8</sup>1990, p.256.

what is now more usually called prospectivism. He says, "...the fact that a course of action would have the best results is not in itself a guide to action, for a guide to action must in some appropriate sense be present to the agent's mind. We need, if you like, a story from the inside of an agent to be part of any theory which is properly a theory in ethics...the passage to action is the very business of ethics."<sup>9</sup> The responsibility constraint is widely embraced by consequentialists.

However, there are other debates where consequentialists have been less willing to accept the application of the responsibility constraint. I think we can understand Bernard Williams's objections to consequentialism as being founded at least in part on the Responsibility Constraint, and consequentialists have, of course, strenuously denied that William's objections are effective. Williams complains that consequentialism is committed to making no distinction between what I do, and what others do or what just happens – Williams calls this the doctrine of negative responsibility, and this is closely related to the doing allowing distinction. As Williams puts it, "From the moral point of view [for the consequentialist], there is no comprehensible difference which consists just in *my bringing about a certain outcome* rather than *someone else's producing it*. That the doctrine of negative responsibility represents in this way the extreme of impartiality, and abstracts from the identity of the agent, leaving just the locus of causal intervention in the world..."<sup>10</sup> Relatedly, Williams' alienation objection can be understood as the complaint that maximisation of utility is an impersonal goal, far removed from the ordinary concerns of ordinary agents, and this is not the sort of thing we *should* be deemed responsible for. And finally, the common worry that consequentialism is too demanding can be understood as relying on the claim that there are limits to what we are reasonably deemed responsible for.

This illustrates the fact that we are not operating with a clear account of responsibility. We have moved from a fairly thin conception of responsibility to a much thicker one. We all agree with the Responsibility Constraint understood thinly. As I said, we all agree that consequentialism cannot tell us to teletransport to save the drowning person. Yet there is substantial disagreement about the doctrine of negative responsibility (the doing allowing distinction), impartiality, and about the demandingness of consequentialism. We have to think about our conception of

---

<sup>9</sup>1991, pp. 466-467

<sup>10</sup> 1973, p.96.

responsibility - how much substance do we get from the Responsibility Constraint? And the methodology for thinking about it it is not obvious. What does it mean to say that we should or should not be held responsible?

One thing it does not mean, as I said at the beginning, is that the consequences of holding responsible would be desirable or undesirable. We need a background account of responsibility that is independent of consequences, and rests, rather, on some quality of the agent. But if we are broadly compatibilist, that is to say, if we think that appeals to metaphysical free will are not helpful here, what we say about agency and responsibility is pretty open. Thus it would be unfair to complain that we have settled the question of what responsibility is, and we are now arguing about normative issues in discussing, for example, the doing allowing distinction. Rather, there is no clear bright line between our accounts of responsibility itself and our 'normative' account of when we should hold people responsible. Instead, there is a scale: at one end we are talking about responsibility in a morally neutral sense, for example when we talk about control. We don't need to bring in normative concerns to agree that if the wind blew you into the path of the trolley, you are not responsible for stopping the trolley – you were not in control at all. At the other end of the scale are cases where we all agree that the agent is in control and acting in full awareness of the situation, and thus responsible in the central sense, but we might disagree over whether she is responsible for the outcomes of her action in a normative sense. For example, there is disagreement over whether (and how) businesses should be held responsible for damage to the environment in various contexts. This is not really a question about whether they *are* responsible, it is a question about what *we* should do in response to their actions. One way to put it is to say that this is a question about *liability*, rather than responsibility. In the middle, between these two extremes, are the cases that I am interested in. Cases where the descriptive and normative questions shade into each other.

It is clear that part of what is going in the background of discussions about alienation, or doing and allowing, is a question about agency, about what sort of thing we can be responsible for. For example, here is Williams on integrity:

It is absurd to demand of such a man, when the sums come in from the utility network which the projects of others have in part determined, that he should just step aside from his own project and decision and

acknowledge the decision which utilitarian calculation requires. It is to alienate him in a real sense from his actions and the source of his action in his own convictions. It is to make him into a channel between the input of everyone's projects, including his own, and an output of optimific decision; but this is to neglect the extent to which *his* projects and *his* decisions have to be seen as the actions and decisions which flow from the projects and attitudes with which he is most closely identified. It is thus, in the most literal sense, an attack on his integrity.<sup>11</sup>

One way to read this passage is to read Williams as saying that we shouldn't demand a utilitarian calculation because it would be *bad* for the agent – it would ask him to do something that would be damaging to him. Thus we might understand the objection as normative – an instance of the usual sort of intuitive counterexample to normative theories: strict deontology is counterintuitive because it would tell you to lie to the enquiring murderer; consequentialism is counterintuitive because it tells you to lynch an innocent person to pacify a mob, and so on. We could understand Williams as saying, consequentialism is counterintuitive because it tells us to neglect our own projects.

On the other hand, we might understand Williams' objection as conceptual. The demand is *absurd* because it violates some conceptual requirement – it is a bit like asking someone to teletransport – this is just not the sort of thing that counts as an exercise of agency. In other words the utilitarian normative theory violates the Responsibility Constraint. Here is Williams again, this time on the example where Pedro (at gunpoint) offers Jim a choice between killing one person himself and 'allowing' Pedro to kill all twenty: "That may be enough for us to speak, in some sense, of Jim's responsibility for that outcome, if it occurs; but it is certainly not enough...for us to speak of Jim's *making* those things happen." *Pedro* is the person who makes the effects."<sup>12</sup> The point here again, is that in the right sense of responsibility, which is (of course) not causal responsibility, Jim is not responsible for the deaths.

I am not endorsing Williams's view of what the Responsibility Constraint says. I think it is controversial – we need to assess it in the context of our best

---

<sup>11</sup> 1973, pp.116-17.

<sup>12</sup> 1973, p.109.

accounts of responsibility. My point here is to shift the argument away from a purely normative argument, to an argument about responsibility. I think the traditional picture neglects the role of accounts of responsibility in thinking about the shape of normative theories. So it is often said, as if it is obvious, that consequentialists cannot countenance a distinction between doing and allowing.<sup>13</sup> And it does seem obvious if we think that the issue is a purely normative one, i.e. a question of what sort of moral principle seems most intuitively attractive in the context of familiar thought experiments. Should you kill one to save five? Consequentialists have looked at the consequences, and assumed that they must come down on the side that results in fewer deaths. And then, to justify this normative conclusion, they have criticised the distinction between doing and allowing on the grounds that it is basically selfish: to think that it is permissible to allow five to die is squeamish – it must be a fear of dirty hands, or excessive concern with one's own acts, or something.<sup>14</sup>

More recently, consequentialists have argued that we can endorse a doing allowing distinction if we allow some agent-centred constraints.<sup>15</sup> This sort of argument accepts the basic framework of the question as a purely normative issue, and abandons impartiality on normative grounds.

Deontologists have similarly accepted that these questions about doing and allowing, impartiality and demandingness are normative questions, and have argued that the doing allowing distinction, or limits on demandingness, can be defended by thinking about some aspect of our normative theory. For example, Philippa Foot and Judith Thomson have both argued that a distinction between positive and negative rights can be used to show that doing harm is worse than allowing harm.<sup>16</sup>

Thinking about these challenges as being rooted in the Responsibility Constraint changes the landscape. Consequentialists usually accept that there is such a constraint, and are happy with the restrictions imposed by it when a very thin conception of responsibility is in play. It is open to consequentialists to accept a thicker conception of responsibility, and thus allow, without having to defend agent

---

<sup>13</sup> See Paul Hurley's chapter in this volume for more discussion of the doing allowing distinction. Hurley accepts that the best consequentialist strategy for accommodating the doing allowing distinction is to admit agent relative elements. Neither Hurley nor I are concerned with the issue of how to draw the line between doings and allowings. For discussion of that see Howard-Synder's SEP entry (2011).

<sup>14</sup> This is an exaggerated version of Philip Pettit's argument (1997).

<sup>15</sup> See Portmore 2011 for a thorough account of how consequentialism can allow agent relativity.

<sup>16</sup> Foot 1994.



relativity for its own sake, restrictions on what can be asked of an agent. As with any normative theory, the consequentialist account of right action must limit itself to the realm of what we can reasonably be deemed responsible for. If I cannot be responsible for unforeseeable consequences, they cannot figure in an account of right action. By extension, if I cannot be responsible for what others do, or what happens, or the impersonal good, then those things cannot be part of what makes actions right.

Of course this raises two difficult questions: first, what does our best account of responsibility say? Do we have any reason at all to think that we can't be responsible in these more controversial cases? And second, if we limited the consequentialist normative theory more dramatically in response to the Responsibility Constraint, do we still have a consequentialist theory? I cannot answer these questions thoroughly in this article, but I will discuss both. I will argue that there is at least one case, coercion, where the Responsibility Constraint applies to consequentialism. I will go on to argue that even if we accept a fairly thick reading of responsibility in the Responsibility Constraint we still have a consequentialist theory.

### 3. Accounts of Responsibility

There are lots of different ways of taxonomising accounts of responsibility, but for my purposes here the most useful distinction is between normative accounts and descriptive accounts. R.J. Wallace provides a thoroughly worked out normative account, according to which an agent should be deemed responsible when it would be fair to do so.<sup>17</sup> This is a normative account because it takes the rules of assigning responsibility to be normative rules. By contrast a descriptive theory of responsibility points to some feature of an agent, and locates responsibility there. Libertarian accounts are descriptivist in this sense, in saying simply (or not so simply) that an agent is responsible when she acts with a free will. Many compatibilist accounts are descriptivist too, even those that draw substantially on Peter Strawson's account.<sup>18</sup> For a compatibilist account to be descriptivist is just for it to say that responsibility depends on some property of the agent, usually something in her quality of will.

---

<sup>17</sup> Wallace 1996.

<sup>18</sup> There are many strands in Strawson's 'Freedom and Resentment': one is the claim that a crucial part of our responsibility practice is the reactive attitudes, but another equally important strand is about the quality of will that provokes those attitudes.

It might seem that normative accounts of responsibility would be no help here. The argument might appear to go as follows: we start with a question about whether there can be a doing allowing distinction, or whether we can be asked to take into account impersonal good, and so on. Now, I am suggesting that one reason we might answer these questions in the negative is that we have a prior commitment to an account of responsibility that limits what we can be asked to do. However, if our account of responsibility is normative, our account of responsibility reaches an answer by asking the exact same question that we started with – should we be held responsible for allowing as well as doing? It would be circular to claim that considerations about responsibility could support the answer to the moral question, they are two sides of the same coin.

This is how we might understand Frances Kamm's discussion of negative responsibility. Kamm argues that in Williams' case, Jim is not responsible for the death of one person if he chooses to kill in response to Pedro's offer. Kamm argues that, rather, the person who makes the offer is responsible.<sup>19</sup> However, as I have argued elsewhere, Kamm's argument does not rely on any independent support from a substantive account of responsibility. Rather, Kamm is examining the normative rules that surround offers and threats. We could, without losing the main thrust of her argument, replace talk of responsibility with talk of right and wrong action: the instigator of the offer (Pedro) has acted wrongly, the person who accepts the offer (Jim) has acted less wrongly, or not wrongly at all depending on small variations in the case.

Samuel Scheffler has an ingenious argument to show that in fact a normative account of responsibility does provide independent support for a doing allowing distinction.<sup>20</sup> Scheffler appeals to a feature of responsibility that goes beyond equivalence with moral rules. Scheffler argues that in seeing oneself as an agent at all, one has to accept that *doing* – a primary exercise of agency – is more important than *allowing*, which is just a secondary exercise of agency. Thus Scheffler thinks that being a responsible agent has a conceptual link to accepting the doing allowing distinction. Scheffler's argument is complex, and I will not go further into the details. I mention it only briefly to draw attention to the form. Scheffler, like Williams, thinks

---

<sup>19</sup> See Kamm 2007. I discuss Kamm's argument in more detail in my 2012.

<sup>20</sup> Scheffler 2004. For criticism of Scheffler's view see Bradley and Stocker 2005.

that denying a doing allowing distinction is problematic because it violates the Responsibility Constraint.

Scheffler's line, that the doing allowing distinction is conceptually mandated by the Responsibility Constraint is very ambitious. It might be easier to show that the doing allowing distinction, or one of the other issues I mentioned, is contingently, as a matter of fact, mandated. With descriptive accounts of responsibility we do not face the danger of circularity. Plausible descriptive accounts of responsibility locate responsibility in an agent's will, focussing on features such as identification with the action (as Frankfurt suggests), or reasons responsiveness (as Fischer and Ravizza, Wolf and others suggest), or some combination of both of those things.<sup>21</sup> It is clearly possible that some particular substantive account would rule out responsibility for allowings, or for the impersonal good, because these things do not bear the right relationship to the agent's will. This, I think, is what Williams is getting at. Williams' thought is that we are so constituted, as a matter of fact, that we cannot have the right sort of relationship to our actions if our actions are guided by consequentialism.

If, however, consequentialists accept that the Responsibility Constraint applies here, they can accept some version of a doing allowing distinction. Here is one argument along these lines.<sup>22</sup> In the case of coercion, where an agent's will is dominated by another agent, as Jim is by Pedro, the coerced is not responsible because they do not bear the right relationship to their action. This is so even though they act (in some sense) voluntarily, and, we can say for the sake of argument, for good reasons. The problem with Jim is that when Pedro has dominated his will, his action is not his own in the right way for him to be responsible for it. To put it in Frankfurt's terms, it is not the sort of action that he can identify with in the right way.<sup>23</sup> This is not a precise account of the problem: obviously there is more to be said. I am just trying to illustrate the fact that accounts of responsibility lend themselves readily to this approach. So even a consequentialist can say that Jim is not responsible when he kills one person in response to Pedro's threat.

The next step is to go back to right action. If Jim is not responsible when he is coerced, what does that say about whether an account of right action can tell him what

---

<sup>21</sup> See Frankfurt 1971 for Frankfurt's original statement of this idea, which is developed in the essays in Frankfurt 1988. Wolf 1990, Fischer and Ravizza 1998, For overviews of recent work in Moral Responsibility see Fischer 1999, Mason 2005, Levy and McKenna 2009.

<sup>22</sup> I give this argument in Mason 2012.

<sup>23</sup> See 'Identification and Wholeheartedness' in Frankfurt 1988.

to do in that situation? The Responsibility Constraint links right action to responsibility – an account of right action must be such that the agent could reasonably be deemed responsible for that action. So if the action is not one that the agent could reasonably be deemed responsible for, it cannot be right or wrong. Does that mean the moral theory has nothing at all to say?

The situation with coercion is rather complicated, as in cases like Jim's the coercer is manipulating Jim's reasons: i.e. the coercer is to some extent taking Jim's point of view in order to know what the reasons that Jim will respond to are. Jim thinks (as most people think) that it would be better for one person to die and nineteen people to live than for twenty to die. Pedro has, without Jim's consent, given Jim a causal role in the outcome. At this point, Jim's agency has been overridden, and so if the Responsibility Constraint applies here, the normative theory cannot tell him what to do. Yet, in some important sense it seems that Jim should shoot, just that when he does, he is not responsible. How can this be? One thing we might mean when we say that Jim is not responsible, is simply that it is permissible for him to shoot. This was how I interpreted Kamm's use of the notion of responsibility here: Kamm is not really saying that Jim has ceased functioning as an agent, she is saying that normative situation is very complex. However, I want to say more, so I face this puzzle – on the one hand Jim should shoot, on the other, he is not functioning as an agent when he does.

Coercion, as I am understanding it in this case, involves one person setting up the reasons so that it is overwhelmingly reasonable for the coercee to do what the coercer wants to get them to do.<sup>24</sup> There may be other sorts of coercion, where the coercer manipulates the coercee using non-rational means, but in this sort of case, the coercion depends on the coercee's responsiveness to reasons. Additionally, in order for the coercer's offer to be coercive, it must make the reasons very powerful. A small balance of reasons in favour of the coercer's preferred option is not coercion. It must be that the coercee would be being unreasonable to refuse the offer – would be doing something that is obviously crazy, to put it more colloquially. So part of my claim, in claiming that the coercee's agency is lost here, is that the agent can be responsive to reasons and yet not fully functioning as an agent. The background account of

---

<sup>24</sup> Note that the situation in coercion is not necessarily that the coercer wants them to *do* it – Pedro might not care either way – Pedro's aim with respect to controlling Jim is not necessarily his overall aim - he wants to *get* Jim to kill someone. The aim is domination.

responsibility goes beyond reasons-responsiveness and involves something like identification as well.

Jim is responding to reasons when he accepts Pedro's offer, in fact, more than that, he is overwhelmed by the reasons. So what sort of reasons are in play here? The puzzle is that if there is no responsibility, there cannot be a directive from the normative theory – just as the normative theory cannot tell an agent to produce the best consequences in cases where the consequences are unforeseeable. This case is similar, in that the agent's actions will affect a state of affairs that clearly has value. Unforeseeable consequences have value, even the consequentialist normative theory, limited by the Responsibility Constraint, must remain silent about what the agent should do (it can't say, 'produce the best consequences', because that violates the Responsibility Constraint). But this case is trickier, because Jim can see the reasons, and in some sense he can act in accordance with them or against them, and so it seems that consequentialism tells him to do what will have the best consequences here just as it does elsewhere.

My view is that consequentialism does not tell Jim what to do in response to Pedro's offer. Jim is put in a position where his agency is compromised, and so the normative theory must bow out, so to speak. But the theory of value remains untouched – it is still the case that it is better for nineteen people to live than for none to live. And this is how the coercion works in this case: Pedro puts Jim in the causal chain. Jim can see that his staying in results in more value, so it would be unreasonable of him to opt out. But he does not bear the right relationship to his action for it to count as responsible.

To see that this is plausible, think about whether we should say that Jim would be acting wrongly if he refused Pedro's offer. It might seem that a consequentialist must say that that would be wrong of Jim. But of course if the consequentialist says that, it seems they are also committed to saying that it would be right to accept the offer. So the consequentialist should not say that it would be wrong to refuse. But notice that this does not commit the consequentialist to saying that it would be *permissible* to refuse the offer. Rather, the consequentialist line must be that the normative theory is silent here. Our theory of value tells us that nineteen lives and one death are much better than twenty deaths – I am assuming for the sake of argument that this is uncontroversial. So the reasons Jim is responding to come directly from the theory of

value, and of course, that is how Pedro has set things up, so that it would be unreasonable of Jim to refuse the offer.

Before I go further I should be explicit about my methodology here. My argument is that there is a distinction between acting wrongly and acting unreasonably, and in the case of Jim, it would be unreasonable but not wrong if he refuses Pedro's offer. This is not supposed to be an argument that makes sense of all our pre-theoretical intuitions about the case. It is a philosophical fix. But I am not perturbed by the somewhat awkward fit with our pre-theoretical intuitions about the case and about the available vocabulary. First, the area of responsibility is one where we are particularly unreliable in our intuitions, as many commentators have pointed out.<sup>25</sup> Second, the terms 'rightness' and 'wrongness' are used in many different ways, and some of those ways are very broad, such that there is no relationship between rightness and wrongness and responsibility – the terms are used as if they are co-extensive with badness and goodness. Inevitably then, a narrower usage of those terms will conflict with some linguistic intuitions. I am proposing a narrow, and in some ways odd, but (I hope) fruitful use for those terms.

My claim again then, is that Jim would be acting unreasonably but not wrongly if he refuses Pedro's offer. He would be acting unreasonably because there are good reasons to accept the offer, namely the fact that nineteen people will live. Because of the situation, these reasons are not rightness and wrongness making reasons. Jim does not bear the right relation to the reasons for them to have that status. But they are nonetheless reasons, and Jim would be unreasonable if he ignored them. This is somewhat analogous to what we might say about reasons of prudence. If we have a view that leaves self-interested action out of the realm of right and wrong, it is clear that someone can act against their reasons without rendering their action wrong. We might further think that reasons of prudence are reasons of value – my welfare matters morally, it provides me with reasons, but it does not provide me with rightness and wrongness making reasons. That's because of the relation I bear to those reasons – according to a fairly plausible view, when the welfare is my own I have a special relationship to the reason producing power of that welfare. Similarly here, the lives of the nineteen people matter, and produce reasons, but in Williams'

---

<sup>25</sup>See particularly Vargas 2013.

example Jim does not have the right sort of relationship to the reason producing power of those nineteen lives for them to be rightness and wrongness making.

I have argued that Jim would be unreasonable but not acting wrongly in refusing Pedro's offer. He might also be morally bad, even though he would not have acted wrongly. One can be morally bad without acting wrongly. One obvious way is through having bad motivations that never result in action. The situation where Jim could justly be accused of moral badness because he refuses the offer is similar in that it depends on Jim's motivations not his action. If Jim refuses the offer through cowardice, he is somewhat bad. If he refuses through a desire for revenge on the local community for some past slight that Pedro knows nothing about, then he is very bad indeed. However, if he refuses because he finds himself constitutionally unable to kill a person, he is not bad at all.

I will make one final remark about this argument. It illustrates one clear superiority of consequentialism over pure deontology. By 'pure deontology' I mean the view that actions are right or wrong simply as actions, and does not base rightness on consequences at all. Imagine a pure deontologist in Jim's situation. Their theory tells them not to kill. It doesn't tell them anything else – a pure deontology (by stipulation) has only a theory of right action, and no theory of the good. So it has no resources for saying anything at all about other ethically relevant features of a situation. What the consequentialist says is, on my picture, very complex. But in this sort of case the simplicity of pure deontology seems like naivety.

In sum, in cases where the agent cannot be responsible for an action, that is not the sort of action that can be right or wrong. That's just what the Responsibility Constraint says, and it is not controversial. I have argued that the Responsibility Constraint may take us into territory that is usually thought of as purely normative, and so may give consequentialism resources to, for example, admit a doing allowing distinction. Of course, coercion is a very complex case, and my argument here depends essentially on the presence of an intervening agent. What should we say about cases where some see a doing allowing distinction and there is no intervening agent? Or the cases I mentioned earlier, where the more general issue of whether we can be deemed responsible for the impersonal good, demanding as that is? I am agnostic: I don't know whether a good argument can be given for thinking that the agent's responsibility is undermined by these demands. There is some reason for pessimism: we are dealing in the grey area between normative intuitions about

holding responsible – or, liability as I called it before, and intuitions about descriptive responsibility. So in considering whether we are responsible in these complex cases, like doing and allowing, we will inevitably bring in our normative intuitions. Thus we are back to the worry that responsibility and rightness are just two sides of the same coin. It seems very likely that consequentialist will tend to see responsibility as permissive, granting that we can be responsible for allowings as well as doings, and for (foreseeable) effects on others, no matter how distant those others are.

#### 4. Cluelessness and Swamping

All this brings us to the worry that the Responsibility Constraint is actually fatal to consequentialism as a normative theory. Consequentialism has two parts, a theory of the good and a theory of the right. The theory of the good has something to say about everything – the consequences we cannot control, the ones we cannot foresee, the ones we are not properly agents with respect to, and so on. But the theory of the right is much more limited, it is limited to those consequences that we can reasonably be deemed responsible for. But now we have a tension. The consequences that are distant still matter, and we are causally related to them, so our normative theory, in directing us only regarding consequences we can be responsible for, has lost sight of its *raison d'être*. This is James Lenman's worry, in his discussion of the epistemic objection. As James Lenman points out, it seems that consequentialists, at least 'pure consequentialists' should be concerned with *all* consequences. And our acts have massive casual ramifications, ramifications, that, given our theory of value, matter as much as anything we can predict. On the one hand consequentialists say that what matters is all these distant states of affairs, on the other hand they say that what we should do is guided by a tiny corner of the vast plain of what matters.

It is worth distinguishing between the original epistemic objection, and Lenman's cluelessness objection. The original objection is that we don't know what the consequences of our actions will be. And the answer to that is straightforward – consequentialists agree that there is a Responsibility Constraint, and so consequentialism concerns only what an agent can be responsible for – in this context let's use Lenman's terminology, and call that 'visible consequences'. But Lenman has a further worry:

A consequentialist might seek to agree to this limiting of moral focus by again taking the line that, in matters of assessing the moral or rational



merits of actions, of assigning praise and blame, we should concern ourselves with subjective rightness, for subjective rightness is precisely concerned only with visible consequences. The trouble is that, as I have noted, a consequentialist must understand this concern as motivated by the belief that maximizing value with respect to visible consequences is a reliable means to maximizing value with respect to overall consequences. And this belief does not appear at all secure. Given this, we might prefer a theory that tells a different story about what the point is of our concern with visible consequences. And such a story would precisely not be consequentialist.<sup>26</sup>

It is important to get the deep point clear here. Lenman grants the indifference postulate for the sake of argument, and we might think that this does solve the problem. If we are ‘entitled’ to the indifference postulate, then all we have to count are visible consequences. But we need to be clear about what we mean when we say that we are entitled to an indifference postulate: if we mean that we are entitled to it because the probabilities really do suggest that the value of the unforeseeable future is neutral, then in granting us an indifference postulate Lenman is granting us the solution to his problem.<sup>27</sup> If, however, we mean that we will use it because we don’t have anything else to go on, then we have solved the epistemic problem, but not the swamping problem.<sup>28</sup> Those consequences are out there, and by consequentialist lights they should be guiding us. They matter, and if consequentialism is a normative theory that makes rightness depends on consequences, it is woefully short of living up to its name.

Lenman says that a consequentialist must think that maximizing value with respect to visible consequences is a reliable means to maximizing value with respect to overall consequences. Obviously the consequentialist should not say this. The reason to maximise the value of visible consequences is *not* that that

---

<sup>26</sup> Lenman 2000, p.365.

<sup>27</sup> Smart, Kagan and Norcross all seem to be answering the epistemic objection in this way.

<sup>28</sup> Lenman op.cit p.357, See also Burch-Brown 2014, p.110 for a very clear statement of this distinction. In my earlier paper on this topic (2004) I argue for an indifference postulate that admits ignorance, but I try to show that there are too many companions in guilt to worry about it. I now see that more needs to be said in response to the swamping objection. One option for consequentialists would be to insist that the right action really is the one that has the best consequences, and argue that the Responsibility Constraint is not important. I do not think that this is a promising strategy.

will entail maximising overall. So what is it? What we don't want to say is that we maximise the value of visible consequences because, given the Responsibility Constraint, that is all we can do. That's exactly the view that makes consequentialism seem lame – it is like saying, 'I want to go to Spain, but it's really too far, so I will just go and hang out at the railway station' – what we ended up doing bears very little relation to what we started out wanting to do.

Here is another way to put the problem: as Lenman points out, all plausible moral theories are concerned with visible consequences. But it seems that the consequentialist rationale for being concerned with only the visible ones is unsatisfactory. We might as well be deontologists, or virtue ethicists, and admit that our concerns are rather short range, that what is close to us is more morally relevant, and so on. Lenman points out a connection between the epistemic objection and Williams' integrity objection: he argues that in allowing too much into the moral picture, we lose track of what a human life is really about – the individual is lost in the massively expanded ethical sphere.<sup>29</sup>

But this looks like an appeal to the Responsibility Constraint, and if it is, then consequentialism and other ethical theories are on very similar ground. The following passage from Lenman makes it look very much as though he thinks that any theory would be concerned with all consequences if that were consistent with the Responsibility Constraint:

What is common ground to all plausible ethical theories is the moral significance of visible consequences. When we can foresee harm to others in the outcome of our actions, we owe them the respect of taking this properly into account. And we owe it to others also to be adequately conscientious in foreseeing such harm. Of course, the invisible consequences of action very plausibly matter too, but there is no clear reason to suppose this mattering to be a matter of moral significance any more than the consequences, visible or otherwise, of earthquakes or meteor impacts (although they may certainly matter enormously) need be matters of, in particular, moral concern.<sup>30</sup>

---

<sup>29</sup> Op. cit. p.370.

<sup>30</sup> Op. cit. pp.364-365.

Notice that this is exactly what I have argued a consequentialist can say: consequences we are not responsible for matter – the theory of the good has something to say about them, but they cannot be a part of normative theory because they are not the sort of thing we can be responsible for. So it seems that both Lenman and the consequentialist want to say the same thing. All consequences matter, but not all can be the subject of normative theory. So it looks as though consequentialism and non-consequentialism are in the same boat with respect to the Responsibility Constraint.

Of course the non-consequentialist could say that consequences we are not responsible for don't matter. But that's a very odd view.<sup>31</sup> For both consequentialism and non-consequentialism, unforeseeable consequences matter. The contrast between the consequentialist and the non-consequentialist lies in the *other* sorts of thing that matter. According to non-consequentialists, consequences matter, but other things matter too – some actions are right or wrong just for their own sake, or because of their relationship to character and virtue rather than to consequences. Thus understood, the problem for the consequentialist rationale in relation to the Responsibility Constraint must be that whereas for non-consequentialist, the Responsibility Constraint just limits one part of the theory, for consequentialism, the Responsibility Constraint limits the whole theory.

But put like this, the swamping problem does not seem like a very serious worry. There are no grounds for thinking that consequentialism is *undermined* by the Responsibility Constraint. It is simply limited by it. But we always knew that – all theories are limited by it. A limit is not necessarily a bad thing, and not necessarily in tension with the thing it limits: a party that must end by midnight is still a party. Consequentialism is a normative theory like any other, and comes with limits.

### Conclusion

There is a Responsibility Constraint on consequentialism, and that might shape the theory more radically than we have hitherto supposed. I have not attempted to say how much, or in what direction the Responsibility Constraint can limit consequentialism, that is work to be done. I have argued that there nothing sinister

---

<sup>31</sup> Pure deontologists don't have to claim that distant consequences don't matter, in that there is no contrast with the consequences that do matter – for pure deontology, morality is just about action.

or undermining about the Responsibility Constraint, we don't have to worry that the rationale for consequentialism is swamped if we accept the constraint. We may end up with a theory that has more in common with non-consequentialist views than traditional versions of consequentialism, because in accepting the Responsibility Constraint we allow fewer consequences into the normative theory. However, such a view is still a distinctively consequentialist view insofar as consequentialism allows *only* consequences into the normative theory.

## **Bibliography**

Bentham, Jeremy 1789. *An Introduction to the Principles of Morals and Legislation*, (J. H. Burns & H.L.A. Hart, eds., Oxford: Oxford University Press, 1996).

Bradley, Ben & Stocker, Michael (2005). “‘Doing and allowing’ and doing and allowing’, *Ethics* 115 (4):799-808.

Burch-Brown, J. M. 2014, 'Clues for Consequentialists', *Utilitas*, 26, 1, pp. 105-119.

Driver, Julia 2011. *Consequentialism*. Routledge.

Feldman, Fred 2012. ‘True and Useful: On the Structure of a Two Level Normative Theory’, *Utilitas* 24(2): 151-71.

Fischer, John Martin 1999. Recent work on moral responsibility. *Ethics* 110 (1):93–139.

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

Foot, Phillipa 1994. ‘Killing and Letting Die’ in *Killing and Letting Die* ed.s Steinbock and Norcross.

Frankfurt, Harry G. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy* 68 (1):5-20.

Frankfurt, Harry G. (1988). *The Importance of What We Care About: Philosophical Essays*. Cambridge University Press.

Gruzalski, Bart 1981. ‘Foreseeable Consequence Utilitarianism’ *Australasian Journal of Philosophy* 59: 163-176.

- Howard Snyder, Frances 1997. 'The Rejection of Objective Consequentialism' *Utilitas* 9 (02):241-248.
- 2011, "Doing vs. Allowing Harm", *The Stanford Encyclopedia of Philosophy* (Winter 2011 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2011/entries/doing-allowing/>.
- Jackson, Frank. 1991. 'Decision-Theoretic Consequentialism and the Nearest and Dearest Objection' *Ethics* 101 (3):461-482.
- Kagan, Shelly 1998. *Normative Ethics* (Boulder: Westview Press)
- Kamm, F. M. (2007). *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*. New York ;Oxford University Press.
- Lenman, James (2000). 'Consequentialism and cluelessness', *Philosophy and Public Affairs* 29 (4):342–370.
- Levy, Neil & McKenna, Michael (2009). Recent work on free will and moral responsibility. *Philosophy Compass* 4 (1):96-133.
- Mason, Elinor 2003. 'Consequentialism and the "Ought Implies Can" Principle', *American Philosophical Quarterly* 40, 319-331.
- 2004. 'Consequentialism and The Principle of Indifference', *Utilitas* 16, 316-321
- 2005. Moral responsibility. *Philosophical Books* 46 (4):343-353.
- 2012. Coercion and Integrity. In Mark Timmons (ed.), *Oxford Studies in Normative Ethics* 2. Oxford University Press, pp.180-205.
- Mill, John Stuart 1863. *Utilitarianism* in *The Collected Works of John Stuart Mill*. Gen. Ed. John M. Robson. 33 vols. Toronto: University of Toronto Press, 1963-91.
- Miller, Dale E. 2003. 'Actual-consequence act utilitarianism and the best possible humans', *Ratio* 16 (1):49–62.

Norcross, Alastair 1990. 'Consequentialism and the Unforeseeable Future', *Analysis* 50, 253–56.

----2006. 'The Scalar Approach to Utilitarianism.,' in Henry West (ed.), *Blackwell Guide to Mill's Utilitarianism* (Oxford University Press), pp. 217-32.

Oddie, Graham and Menzies, Peter 1992. 'An Objectivist's Guide to Subjective Value', *Ethics* 102: 512-533.

Pettit, Philip 1997. 'The Consequentialist Perspective' in M. Baron et al. *Three Methods of Ethics; A Debate*, Oxford: Blackwell) pp. 92-174.

Portmore, Douglas W. 2011. *Commonsense Consequentialism: Wherein Morality Meets Rationality*. Oxford University Press.

Scheffler, Samuel 2004. 'Doing and Allowing', *Ethics* 114 (2):215-239.

Smart, J. J. C. 1963. Free will, praise and blame. *Mind* 70 (279):291-306.

---- 'An Outline of a System of Utilitarian Ethics' in *Utilitarianism: For and Against*, by J.J.C. Smart and B. Williams. Cambridge: Cambridge University Press. p. 3 – 74.

Smith, Holly 2010. 'Subjective Rightness' *Social Philosophy and Policy* 27 (2): 64-110.

Steinbock, Bonnie, and Norcross, Alastair (eds.), 1994, *Killing and Letting Die*, 2nd edition, New York: Fordham University Press.

Strawson, Peter F. (1962). Freedom and resentment. *Proceedings of the British Academy* 48:1-25.

Thomson, Judith Jarvis, 1986, "Killing, Letting Die and the Trolley Problem," *Rights, Restitution, and Risk: Essays in Moral Theory*, W. Parent (ed.), Cambridge: Harvard University Press.

Timmons, Mark. (2002): *Moral Theory: An Introduction*, New York and Oxford: Rowman & Littlefield.

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

Wallace, R. Jay 1996. *Responsibility and the Moral Sentiments*. Harvard University Press.

Wiland, Eric 2005. 'Monkeys, typewriters, and objective consequentialism', *Ratio* 18 (3):352–360.

Williams, Bernard, 1973, "A Critique of Utilitarianism," in *Utilitarianism: For and Against*, by J.J.C. Smart and B. Williams. Cambridge: Cambridge University Press, 77-155.

Wolf, Susan 1990. *Freedom Within Reason* Oxford University Press.

Zimmerman, Michael J. 2008. *Living with Uncertainty* Cambridge: Cambridge University Press.