

What Pedro Could Do*

Christopher Woodard

In Bernard Williams's famous story, Jim must choose whether to shoot an innocent hostage. If he does not, Pedro will shoot that person plus nineteen more. If Jim does shoot, Pedro will release the other nineteen hostages. Jim must decide whether to do something terrible. If he does not, these innocent people will bear an enormous cost.¹

The main point of Williams's discussion is not about whether Jim should shoot—he allows that, perhaps, he should—but instead about what Jim's reasons are. Williams supposes that, whatever the verdict about what Jim should do, Jim certainly has a strong reason not to shoot. This, he thinks, is sufficient to show that Act Utilitarianism is strongly counter-intuitive, since Act Utilitarianism apparently cannot account for this reason.

Suppose that Williams is right that Jim has a strong reason not to shoot. Let us add, as seems undeniable, that Jim has a strong reason *to* shoot—since doing so would save nineteen innocent lives.² Let us also shelve the question of what Jim should do, all things considered. Which sort of ethical theory seems best placed to explain the existence of these countervailing reasons?

This question is importantly broader than the one that Williams and most of his commentators went on to discuss. Their question was about how to account for Jim's strong reason not to shoot—whether in terms of integrity, or agent-relativity, for example, or something else. The broader question is how best to account for *both* of Jim's reasons: to shoot, and not to shoot. Ideally, we would like a satisfying ethical theory to explain the strong conviction that there are conflicting reasons in cases such as this. These are cases in which failing to do the thing that we are certain is morally wrong in ordinary cases has a very high cost. A theory which explains one of Jim's reasons without explaining the other

* I am grateful to Stephen Barker, David Beesley, Gregory Currie, Guy Fletcher, the late Susan Hurley, Mark Jago, Gregory Mason, Douglas Portmore, and members of an audience at the University of Nottingham for very helpful comments on earlier versions of this paper. © Christopher Woodard

¹ See Williams (1973: 98-100).

² For simplicity, assume that any shooting will certainly kill the victim.

would, I take it, be inferior to another which explained both of Jim's reasons, other things equal.

In this paper I'll contrast two broad approaches to explaining conflicting reasons in cases like this. One, which has received most attention in recent years, focuses on the agent faced with the dilemma, who I'll call *the actor*. For example, this approach focuses on Jim. Advocates of it usually hope to find a reason not to shoot in his integrity, or his special relationship to his own actions, or his distinctive point of view on the world. According to this sort of view, we might say, Act Utilitarianism goes wrong by failing to recognise the full significance of Jim's agency.

I'll favour a different approach. This focuses on the agent who sets up the dilemma: the one who creates the obstacle that forces the actor to choose between what seems right and what seems good. I'll call this individual *the other agent*. For example, this approach focuses on Pedro. According to this sort of view, Act Utilitarianism goes wrong by failing to pay due attention to what Pedro could do, as contrasted with what he will or would do. It fails to recognise the full significance of *Pedro's* agency.

I'll begin by outlining the two different approaches. Then I'll explain the attractions of claiming that Jim's reasons depend on facts about what Pedro could do. Of course, other approaches are possible. Hence, my conclusion is only comparative: the analysis of Jim's reasons in terms of what Pedro could do is at least as promising as the dominant approach, which analyses Jim's reasons in terms of the special importance of what Jim does.

1. What Jim does

Most responses to Williams's story assume that it suggests that Act Utilitarianism goes wrong by failing to appreciate the special importance of some feature of Jim's agency. Williams himself drew this conclusion from the story. Having just introduced Jim, Williams outlines his diagnosis of the faults of Act Utilitarianism as follows:

A feature of utilitarianism is that it cuts out a kind of consideration which for some others makes a difference to what they feel about such cases: a consideration involving the idea, as we might first and very simply put it, that each of us is specially responsible for what *he* does, rather than for what other people do.³

Famously, Williams goes on to develop this idea through a discussion of integrity, alleging that Act Utilitarianism is committed to an incoherent picture of the relationship between an agent's 'projects and his actions'.⁴

However, let's abstract from the details of the subsequent discussion, and concentrate instead on the underlying simple idea.⁵ Williams is right that a tempting reaction to the case of Jim, and others like it, is that Act Utilitarianism fails to give due weight to the special responsibility each agent has for her own actions. In particular, this is a tempting diagnosis of its failure to explain Jim's strong reason not to shoot. According to the Act Utilitarian, Jim has no reason not to shoot, since his reasons depend only on the consequences for utility of his options in the case at hand. These consequences are as follows: if he shoots, one innocent person will die; if he does not shoot, that innocent person will *still* be shot, and so will another nineteen. Jim has no other options, and there is no uncertainty about these consequences.⁶ However, this bald reckoning seems to ignore a crucial feature of the situation. If Jim shoots, he has killed one innocent person himself; while if he does not, it will be Pedro who kills the twenty. Though these twenty deaths will be a sure consequence of Jim's refusal, they still will not be something that he does. Since he has a special responsibility for what he does, he has a strong reason not to shoot that Act Utilitarianism does not recognise.

At risk of some oversimplification, I will say that diagnoses of this sort—those resting on the idea that what Act Utilitarianism misses in cases like this is the special responsibility of each agent for her own actions—focus attention on what the actor *does*. In Jim's case, they focus attention on *what Jim does*. The idea

³ Williams (1973: 99), emphasis in the original.

⁴ Williams (1973: 100-118). The quoted phrase appears on p. 100.

⁵ For discussion of Williams's claims about integrity, see for example Hollis (1983) and Scheffler (1982: Chapter One).

⁶ Williams's story is slightly more complex, in two main ways. First, there are three agents: Jim, Pedro, and the captain (Pedro's boss). Second, he allows Jim a third option, of trying to overpower Pedro and the captain, the certain result of which would be the death of all twenty hostages plus Jim. I will ignore these features of the case, since they do not bear on the issues at stake here.

is that paying attention to the nature of what he does brings to light important ethical considerations that Act Utilitarianism misses. This risks oversimplification because we can cash out the idea of special responsibility for one's own actions in different ways. For example, we can cash it out in terms of the distinction between doing and allowing, or instead in terms of the distinction between intending and foreseeing. Both of these distinctions are, from an abstract point of view, versions of the same idea, of attaching special importance to the actor's own actions; but, of course, there are other very important differences between these different versions of that idea.

Some version of the idea that agents are specially responsible for their own actions seems embedded in the idea of agent-relative constraints.⁷ Such constraints specify certain kinds of action, and rule out one kind of putative justification for performing acts of these kinds: namely, that doing so is necessary and sufficient to prevent a greater number of acts of the very same kind. For example, we might claim that there is an agent-relative constraint against killing the innocent. This implies that it would not be right for Jim to shoot one innocent person just because doing so is necessary and sufficient (in the circumstances) to prevent Pedro killing that person plus another nineteen. Such a constraint presupposes that we distinguish between what Jim does and the total effects of his actions, for the latter would include Pedro's shootings were Jim to refuse to shoot. So, the idea of such a constraint relies on some such distinction between what agents do and what happens as a result of what they do.⁸ This illustrates just one way we might rely on the idea of special responsibility in thinking about cases like Jim's.

Samuel Scheffler has argued recently that, whether we believe in agent-relative constraints or not, we can't avoid attributing basic ethical significance to some version of the idea that agents are specially responsible for what they do.⁹ Moreover, he makes some intriguing remarks that seem to suggest that some

⁷ See Kagan (1989: 24-32; Chapters 3-4).

⁸ Agent-relative constraints needn't rely on the distinction between doing and allowing in particular; they could, for example, rely on the distinction between intended and merely foreseen consequences of actions. The point is simply that some such distinction—in a broad sense between what the actor is and is not specially responsible for, to use Williams's phrase—is necessary for the idea of an agent-relative constraint. See Kagan (1989: Chapters 3-4).

⁹ Scheffler (2004).

version of this idea enables us to strike a balance between two views of ourselves. For example, he writes:

. . . one of the functions of norms that rely on the distinction between doing and allowing is to regulate the effectively ineliminable tension between the special importance we attach to our own actions, and our recognition that those actions are nevertheless subsumed within the larger causal web.¹⁰

One way of interpreting this claim is as follows. In placing special emphasis on what agents do, norms that distinguish between doing and allowing give due weight to our view of ourselves as responsible agents. On the other hand, in nevertheless taking some account of what agents allow, these norms give due weight to our view of ourselves as mere parts of the ‘causal web’. Moreover, we can reflect our considered views about the relative importance of these two aspects of agents in the norms themselves, by adjusting the extra weight given in those norms to what agents do over what they allow. In this way, we rely on these norms to strike an appropriate balance between these two views of ourselves—as Scheffler puts it, to ‘regulate the effectively ineliminable tension’ between these two ways of thinking of agents.

Does Scheffler’s remark offer the key to explaining both of Jim’s reasons? We might hope so, since we might expect that his reason to shoot is associated with the view of him as part of the causal web, while his reason to not shoot is associated with the view of him as a responsible agent. However, there are some well-known and serious problems with this way of analysing such cases. Briefly, they are as follows.

The first is to explain how to distinguish between what the agent is and is not specially responsible for—between, for example, what he does and what he allows, or between what he intends and what he merely foresees. In some hard cases it is very difficult to draw these distinctions.¹¹ Second, there is a problem in showing why such distinctions, once drawn, are ethically significant—except by

¹⁰ Scheffler (2004: 225). Similarly, he says (2004: 239): ‘part of the function of norms that distinguish between doing and allowing is to strike a balance between considerations of two very different kinds: between the inevitable priority and distinctiveness of action, on the one hand, and its unavoidable subsumption within a larger causal web, on the other’.

¹¹ For discussion of these difficulties, see for example Quinn (1989); Kamm (1992); Bennett (1995: Chapters 4-11); Williams (1995).

appeal to those intuitions we wish to explain in light of one of these distinctions. If that is the most we can say for them, though, we cannot explain these intuitions in terms of either distinction. At best, the distinctions would help us to summarise the convictions. To use them to explain the convictions, we would need some independent rationale for attributing significance to one or both distinctions.¹²

Third, there is what we can call *Kagan's question*: suppose that we have reason to believe that the distinction between doing and allowing (say) has basic ethical significance; why think it has different significance as applied to the actor's actions than as applied to other agents' actions? Suppose, for example, that we employ this distinction in thinking about Jim's case. Why think that what Jim does is more important for Jim's reasons than what Jim allows, but not that what *Pedro* does is more important for Jim's reasons than what *Pedro* allows? On the face of it, applying the distinction to all relevant agents will undermine the attempt to explain common convictions in cases like this; but as Kagan argues, there is no clear rationale for applying it to some but not all relevant agents.¹³

One further problem is that, contrary to what Scheffler seems to suggest, we cannot explain the *conflict* in Jim's reasons merely by appeal to some suitably drawn distinction between what he is and is not specially responsible for. For example, appeal to the distinction between doing and allowing doesn't suffice to explain this conflict. Suppose we think that what is done is twice as important as what is allowed. In considering Jim's reasons we then discount the deaths Pedro would bring about were Jim to refuse to shoot by a factor of two, as compared with the death Jim would bring about were he to shoot. Still we have the conclusion that Jim has a strong reason to shoot, and no reason not to do so. Now suppose that we discount what is merely allowed by a greater factor, sufficient to overturn this judgement about Jim's case (we might say, for example, that what is done is one hundred times more important than what is allowed). Then we would have to conclude that Jim has a strong reason not to shoot, and no reason to shoot. In short, however we balance the significance of doing and allowing, no simple appeal to the distinction can explain what we set out to explain, which is the conflict in Jim's reasons.

¹² Kagan makes this point very forcefully (1989: 114-121).

¹³ See Kagan (1989: 125-126).

Perhaps this last problem is not very severe. After all, it might be objected, this merely shows that a satisfying ethical view would have to contain other components. Admittedly, it is worth pointing out that ‘norms that rely on the distinction between doing and allowing’ don’t suffice to explain the conflict of reasons in Jim’s case and others like it. But it may still be true that they play an important or even essential role in a larger story which does explain such conflicts.

Consider, then, one final problem for approaches which focus on what the actor does. Given their focus, such approaches seem insensitive to the issue of whether an agent or a non-agent causes (or constitutes) the obstacle facing the actor; yet, our intuitions seem sensitive to this issue. In Jim’s case, Pedro stands in the way of Jim and the ideal outcome, in which no one is shot. Pedro’s power and intentions force the unwelcome choice on Jim. However, suppose instead that the obstacle is a causally similar non-agent: a trolley, say. This trolley is in some state such that it will kill all twenty unless Jim first shoots one, in which case the trolley will kill none.¹⁴ Williams famously remarked that there is something wrong in Act Utilitarianism’s requiring the actor to give up ‘a project or attitude round which he has built his life, just because someone else’s projects have so structured the causal scene that that is how the utilitarian sum comes out’.¹⁵ But a non-agent could structure the causal scene in the same way, presenting the actor with a similar unwelcome choice.¹⁶ As Williams’s remark suggests, that sort of example wouldn’t serve his purposes equally well, because our convictions seem sensitive to the issue of whether the obstacle is presented by another agent or not.

If it’s true that our convictions are sensitive to this issue, approaches which focus on the special responsibility of agents for their own actions may be looking in the wrong place. On the face of it, Jim’s responsibility for his own actions does not depend on whether he faces Pedro or a causally similar runaway trolley.

¹⁴ No agent even set the trolley in motion: assume the wind did so. (I thank Guy Fletcher for raising this point.)

¹⁵ Williams (1973: 116).

¹⁶ Scheffler (1982: 10) makes a similar point.

2. What Pedro could do

Instead of focusing on Jim, we could focus on Pedro.¹⁷ Is there some way in which Act Utilitarianism fails to recognise the full significance of the fact that *he* is an agent?

Consider again the comparison between Pedro and a trolley. Why would the trolley not serve Williams's purpose just as well? One possible answer is that Pedro could respond to Jim in a way he will not. That is, there is a contrast between what Pedro *could* do, and what he *would* do under any given conditions. No equivalent contrast applies to the trolley.

At least, this is what commonsense tells us, and what much of ethics takes for granted. According to our commonsense use of the concept of 'could do', it applies to agents in a way that it fails to apply to non-agents. Of course, we use the concept for non-agents too. We use it to signal uncertainty about how things will turn out, or to make claims about what would happen given certain conditions. 'It could rain tomorrow' is an example of signalling uncertainty, while 'this car could do one hundred' is an example of a claim about what would happen given certain conditions—in this case, were someone to try to drive the car at one hundred miles per hour. Sometimes we might use 'could' in a way that mixes these functions, as when someone says, 'that bridge could collapse in a flood'. This claim might signal some uncertainty as well as being about what would happen were a flood to come.

The use we are interested in has nothing to do with uncertainty, so let us set that firmly aside. Instead we are interested in the sense of 'could do' in which we say that agents could perform any one of their options. This seems to be a species of the other main use of 'could do', in which we use it to make claims about what would happen given certain conditions. With respect to agents only,

¹⁷ Frances Kamm also criticises the focus on the actor in cases like this. However, she advocates focusing on *victims*, rather than on what I've called the other agent. She writes of her view (Kamm 2006: 29): 'Unlike an agent-focused account, this explanation does not focus on what I do rather than what others do. The fact that if I kill someone, I would be acting now and the victim would be mine does not play a pivotal role in explaining why I must not kill him even though my duty is an agent-relative duty. We explain why I must not kill him by focusing on each person's inviolability. His right, not my agency, constitutes the moral constraint'. Note that her view seems to require that there is a right-holder on the scene. For that reason, we might doubt whether it explains some intuitions in cases where the harm to be done seems impersonal, as when another agent would destroy every artwork in the Uffizi unless I destroy one.

we use ‘could do’ to highlight what would happen were the agent to have a certain motivation or will. For example, we might say ‘she could train to become a doctor’, meaning that this is one of her options. It seems that this means, roughly, that were she to have the will to train to be a doctor, she would train to be a doctor. This has the same general form as the claim about the car: such and such would happen, were such and such conditions met; it’s just that the conditions in question centrally involve the agent’s will. Note that there need be no question of uncertainty here: we may know for sure that she will not train to be a doctor, because we know for sure that she will not have the will to do so. The claim instead is that, were she to have the will, she would train to be a doctor. Her will is the only relevant obstacle.

Thus, this agent-specific use of ‘could do’ asserts the salience of alternative courses of action, by implicitly or explicitly emphasising the importance of the agent’s will as an enabler or disabler of action in the context at hand.¹⁸ It is as if we were saying, ‘only the agent’s will stands between her and these possible actions’. For example, if I say, ‘I could hold my breath for three minutes’, I am claiming falsely that only my decision not to try stands between me and this feat. We might contrast Pedro and a trolley using just this sense of ‘could do’. While he *could* respond to Jim’s decision in ways that he *would not*, the trolley could not respond in any way other than it would. It’s not that there is greater uncertainty about his response: as Williams constructs the story, we are in no doubt about Pedro’s response to each of Jim’s possible choices. Instead, it is that alternate possibilities seem salient when we consider Pedro’s response, because only his will stands between their possibility and actuality, whilst nothing equivalent applies to the trolley. This is one of the main differences we might have in mind if we point out that Pedro is an agent but the trolley is not.

Should ethical views be sensitive to this particular contrast between agents and non-agents? It’s not immediately clear. This is not the only contrast between agents and non-agents that we could draw; so it’s not as if being blind to this contrast entails treating agents altogether in just the same way as non-agents. We need some positive argument for thinking that some mistake is involved in failing to discriminate in just this way between Pedro and a causally similar trolley.

¹⁸ I shall not try to analyse the concept of an agent’s will. I note only that the ubiquitous concept of an agent’s *options* relies on it.

I propose to proceed in two stages. First, let's see whether caring about the contrast between what Pedro could and would do would enable us to say attractive things about Jim's reasons. If, as I shall claim, it does, there is then a further question as to whether we can give some rationale for caring about this contrast, other than that caring about it has attractive implications. I'll turn to the second question in the following section.

First, then, consider how we might try to explain Jim's reasons in terms of what Pedro could do. Act Utilitarianism explains the actor's reasons in terms of the states of affairs that would result from her action, were she to perform it. Where these results depend on the behaviour of other agents, it examines the way they would behave were the actor to perform the option in question. Now suppose that we seek to explain the actor's reasons in terms of the responses that agents in her environment *could* perform. It's not immediately clear how to do this, for the range of ways those actors could respond may be very wide. Moreover, the possible responses may well be inconsistent with each other (as when, for example, the other agent could either *agree* or *not agree* to the actor's proposal). If we look at the full range of possible responses, we may just have an incoherent picture of the value of each of the actor's options.

One way to avoid this incoherence is to try to explain the actor's reasons in terms of the *best possible* response by other agents. This associates reasons not with the full range of things each other agent could do, and not with the way each other agent would respond, but with the best way they each could respond. For example, we could claim that Jim's reasons depend on the best way that Pedro could respond to each of Jim's options. If Jim shoots, the best way Pedro could respond (by our lights) is to shoot none. The result of this combination of actions is one innocent killed. On the other hand, if Jim does not shoot, Pedro's best response is again to shoot none; but the result of this combination of actions is none killed. If reasons were associated with the best possible response, in this way, Jim would seem to have a strong reason not to shoot.

Call this the 'proto-explanation' of Jim's reason not to shoot. It strongly resembles the claims made by so-called *possibilists* in deontic logic.¹⁹ Possibilists claim that, when considering the deontic status of the actor's options (whether

¹⁹ For the debate between 'actualists' and 'possibilists', see for example: Bergström (1966); Jackson and Pargetter (1986); Jackson (1987); Zimmerman (1996: Chapter 6).

these options are forbidden, required, or optional), we should consider the actor's own best possible response to each of these options. For example, suppose you are asked now whether you will help with some good cause next weekend. Your options now are to commit or not; next weekend you will face a choice whether to help or not. Suppose also that, as a matter of fact, you would not help next weekend even if you commit now, even though it is true that you could help were you to have the requisite will. The problem is not what you could do next weekend, but what you would do, were you to commit now. It's not that helping would be too onerous, or would require an unreasonable degree of sacrifice; it's just that you would lack motivation to do the best you could do. Let's say that the best outcome would result from your committing-and-helping, the next best would result from your not-committing, and the worst would result from your committing-and-not-helping. Should you commit?²⁰

Actualists say not. They claim that the rightness of actions depends exclusively on their 'actual' consequences—where that means, somewhat idiosyncratically, the consequences they *would* have were they performed. In our case, this means that the rightness of your committing depends on the consequences that would result were you to commit, which are worse than the consequences that would result were you not to commit. Possibilists, in contrast, claim that the rightness of actions can depend on 'merely possible' consequences, where these would result were the actor herself to respond in ways that she could. In particular, they instruct us to look at the consequences of each option supposing the actor were to react in the best way she could, even if she would not respond in that way. In our example, that means that it is right to commit, since doing so would have better consequences than any alternative, on the assumption that the actor—you—were to respond in the best way you could, by helping next weekend.

Our proto-explanation of Jim's reasons is similar. It evaluates Jim's options according to the results they would have were relevant agents to respond in the best way they could. In particular, it evaluates Jim's option of not shooting according to the results it would have were Pedro to respond in the best way he could. This option has better consequences (zero deaths) than Jim's other option,

²⁰ This example is adapted from Jackson and Pargetter's case of Professor Procrastinate. See Jackson and Pargetter (1986: 235). See also Jackson (1987: 110, n. 13).

which would result in one death even if all relevant agents respond in the best way they could. There are, though, two important differences between this and the claims that possibilists make. One is that possibilists consider best responses only by the actor herself; for all other agents they consider what would occur, just as the actualist does. Second, possibilists make claims about what agents ought to do, whereas we are considering a claim about Jim's reasons. What ought to be done is downstream of reasons: judgements about what ought to be done are verdicts about the interaction of all the reasons there are for or against the actor's options.²¹ I'll return to the significance of this point in section 4.

Let me summarise the argument so far. For each of Jim's options, we can contrast how Pedro could respond with how he would respond to that option. If we focus on the *best* way Pedro *could* respond, we have a proto-explanation of Jim's reason not to shoot. This explanation resembles the claims made in deontic logic by possibilists, except that it considers the best response by an agent other than the actor, and that it involves claims about reasons not obligations.

Does the similarity with possibilists' claims support the proto-explanation of Jim's reason not to shoot? Unfortunately not, for two reasons. The first is that possibilists have failed to explain why we should consider the actor's best possible response except to say that doing so has attractive implications in some cases. When pressed, they tend to say that failure to consider the actor's best response would let her off the hook too easily, since it releases her from obligations she would have were she a morally better person.²² However, that merely reports our convictions about cases. It does nothing to explain why reasons should depend on 'merely possible' consequences in the way possibilists claim. The situation is comparable to that facing someone who attributes great ethical significance to the distinction between doing and allowing. Without some rationale for the theoretical claim, we cannot explain the intuitions in terms of the theory; at best, the theory would merely summarise the intuitions.²³ Moreover, since no one has identified a rationale for possibilism, we cannot judge whether any rationale it

²¹ See Dancy (2004: 16 n. 2). For an example of the view that obligatoriness sometimes provides an independent reason, and so is not a mere verdict, see Scanlon (1998: 11).

²² See Zimmerman (1996: 193-5, 203-6).

²³ See Kagan (1989: 13-14); Parfit (2007: 264).

might have could extend to cover appeals to the best response by agents other than the actor.

Second, possibilism seems wrongly to ignore the importance of how relevant agents *would* respond. In our earlier example, it seems wrong to ignore the fact that you would not help next weekend, even if you commit now. Surely, we might think, that fact must bear on your reasons and the rightness of your committing or not. The same point applies to Jim's case, of course. We seem to go wrong if we ignore the fact that Pedro would not respond to Jim's refusal to shoot in the best way he could. This fact about Pedro is amongst the most important that Jim must reckon with. Any satisfying ethical view must treat it as significant. As I said at the start, we want an ethical view to explain why Jim has a strong reason to shoot as well as a strong reason not to do so.

For these reasons, the similarity between the proto-explanation and possibilism fails to inspire confidence in the proto-explanation. Can we solve these two problems?

3. Why should we care about what Pedro could do?

The first problem is to find some rationale for caring about what Pedro could do, other than that doing so has implications that seem to match common convictions in this case. In particular, of course, we're looking for some rationale for caring about the best that Pedro could do in response to Jim's options. The idea is that Pedro's possible response, of killing none, to Jim's refusal to shoot, may give Jim a reason to refuse to shoot. How could that be?

Compare this suggestion with the sort of reasons that Act Utilitarianism recognises. According to Act Utilitarianism, the reasons for or against Jim's refusing to shoot have to do with the consequences that refusal would have, as compared with the consequences of his other option in the circumstances, shooting. The possible outcome in which none are shot just doesn't feature in this analysis. Since there is no uncertainty about Pedro's response, this possible response on his part is utterly irrelevant to Jim's reasons, according to Act Utilitarianism.

Act Utilitarianism has a good explanation of this. It claims that reasons for or against an action depend on what that action would cause, in the circumstances. It finds reasons in causal relations between actions and valuable or disvaluable outcomes. Since Jim's refusal to shoot has zero chance of causing the outcome in which none are shot, in the circumstances—because of the facts about how Pedro would respond—this outcome does not bear on Jim's reasons in this case.

This explanation points to an intelligible and appealing theoretical claim about the grounds of reasons for action: such reasons depend on the causal properties of actions, and in particular the good or bad outcomes to which actions are causally related. This is attractive, I take it, because actions are interventions in the causal order; the whole point of acting, it is tempting to believe, is to influence what happens. If so, the claim that causal relations are the grounds of reasons for action has some credibility that is independent of any implications it may have in particular cases. Possibilism fails to underwrite the proto-explanation because it has no equivalent story about the grounds of reasons. If we ask possibilists why counterfactual causal relations between actions and valued outcomes should give rise to reasons, they have no answer other than that supposing they do generates intuitive implications in some cases.

However, there is a possible rationale for possibilists' claims. The key is to think of the valued outcomes in question—in our example, the outcome in which no one is shot—not as *merely possible* outcomes of *individual* actions, but as *actual* outcomes of *patterns of action*. The outcome in which no one is shot is what would occur were Jim and Pedro both not to shoot anyone. This is the actual outcome of that pattern of action, in just the sense at issue between actualists and possibilists: it is what would happen were the action in question performed. It's just that the action in question now is not an individual action, or indeed any action performable by Jim alone. It is a pattern of action, in the minimal sense of a conjunction of actions performable by relevant individual agents. There is an ordinary causal relationship between this pattern and the valued outcome: if performed, the pattern would realise this outcome.

So, rather than look for unusual counterfactual causal relations between individual actions and the valued outcomes, we might explore the idea that the

reasons for or against an action could depend on its being *part* of some pattern of action that is causally related to the valued outcomes. Such reasons would be *pattern-based*:

Pattern-based reasons are reasons for or against some action, *A*, because of the goodness or rightness of some pattern of action, *P*, of which *A* is a part.²⁴

This is the bare idea of pattern-based reasons. Obviously, it falls short of being a theory of such reasons. Most importantly, any theory of pattern-based reasons would have to tell us which patterns of action support reasons, and why. That is, it would have to tell us about the eligibility of patterns.

Nevertheless, even the bare idea should encourage us to take pattern-based reasons seriously. For this idea allows us to make sense of the proto-explanation of Jim's reason not to shoot, and more. Why think that the outcome in which no one is shot bears on Jim's reasons for or against refusing to shoot? The answer is that his refusal is part of the best pattern of action that he and Pedro could perform. On this view, parthood relations are capable of generating reasons, just as causal relations are. Jim's refusal is part of the best pattern he and Pedro could perform, and its being part of this pattern gives Jim a reason to perform it.

Earlier I said that we can understand Act Utilitarianism as embodying an attractive theoretical claim about the grounds of reasons. It embodies the view that causal relations ground reasons, and that the point of acting is to affect what happens. The idea of pattern-based reasons embodies its own claim about the grounds of reasons, which is that parthood relations ground reasons. If that is right, we might say that part of the point of acting is to realise parts of favoured patterns of action, or to make impossible disfavoured patterns of action by refusing to realise one or more of their essential parts.

In fact this idea has some currency in ethics. Collective Consequentialism is a theory of pattern-based reasons: it claims that you should play your part in

²⁴ Here I develop suggestions made by Hurley (1989: Chapter 8); Jackson (1987: 106); and Regan (1980). See also Woodard (2003; 2007; 2008). Where *P* is identical to *A* we have a limiting case of pattern-based reasons. These are, in fact, ordinary *act-based* reasons, according to which the reasons for or against *A* depend on the rightness or goodness of *A* itself. In this way, the idea of pattern-based reasons is more general than the idea of act-based reasons.

the best pattern of action performable by your group, because it is your part in this best pattern.²⁵ Some forms of Rule Consequentialism also are best understood as theories of pattern-based reasons.²⁶ The idea of complicity seems to trade on the suggestion that one should not play one's part in disfavoured patterns, even if doing so is instrumentally neutral or positive when considered as an individual action. The Kantian idea of imperfect duties also seems to embody commitment to pattern-based reasons. Such duties sort lives, or anyway extended portions of lives, rather than individual actions. They say, for example, that a *life* in which one does nothing to develop one's talents is forbidden for that reason. If imperfect duties are ever to give us reasons for or against performing individual actions, it must be because such actions are related by parthood to one of these forbidden or required lives.

Acting with concern for the parthood properties of one's actions seems intelligible. If it makes sense, we can give an independent rationale for the suggestion that the outcome in which no one is shot gives Jim a reason to refuse to shoot. His refusal is, of course, part of the pattern of action performable by him and Pedro that is necessary and sufficient to cause this outcome. If he shoots, he makes this outcome impossible. If parthood properties matter, and the pattern involving neither Jim nor Pedro shooting anyone is eligible, then Jim has a pattern-based reason not to shoot. The plausibility of this explanation depends centrally on whether we can give some account of the eligibility of patterns that is both plausible and implies that the pattern in which neither Jim nor Pedro shoots anyone is eligible to provide Jim with a reason. I shall not attempt to give a full account of eligibility here.²⁷ However, it is worth pointing out that to give this explanation of Jim's reason not to shoot we have to deny one possible constraint on eligibility that has seemed attractive to many. This constraint is the *Willingness Requirement*:

WR There is a pattern-based reason to play one's part *A* in a pattern *P* performable by a group *G* only if the members of *G* are willing to perform *P*.

²⁵ See Parfit (1987: 30-1); Mulgan (2001: Part Two).

²⁶ See Woodard (2008).

²⁷ I discuss issues of eligibility in Woodard (2007: Chapters 4-7).

Different versions of this requirement are possible.²⁸ However, they all express a simple idea: there are no pattern-based reasons in uncooperative contexts.

This idea may seem attractive for various reasons. Note, though, that it rules out concern with complicity (since, typically, issues of complicity arise when the other agents would not be willing to play their parts in the favoured pattern), and it rules out our explanation of Jim's reason not to shoot, since Pedro is unwilling to cooperate in producing the best pattern.

Why should someone inclined to accept the existence of pattern-based reasons think that they could not exist absent willingness on the part of the other agents concerned? There are several possible reasons. One has to do with worrying that we won't be able to find sufficient grounds, other than willingness, to exclude clearly irrelevant patterns. Call this the worry about mere agglomerations. I shall not say much about this here, except to point out that the issue of which patterns are eligible, for someone who believes in pattern-based reasons, resembles the issue about which possible actions count as options, which is an issue for everyone. There may be resources in that resemblance for excluding mere agglomerations on grounds other than facts about willingness.²⁹

There is another worry quite separate from that about mere agglomerations, however. If we deny the Willingness Requirement, how can we give due weight to the importance of facts about how others would respond to the actor's choices? This takes us back to the second problem we identified with the possibilist proposal.

4. Caring about what Pedro could *and* would do

The worry is that if we appeal to the idea of pattern-based reasons in uncooperative contexts, such as Jim's, we will fail to be properly concerned with how other agents would in fact respond. For example, there might be some merit in all UK drivers driving on the right, since this would decrease the risk of

²⁸ See Woodard (2007: 71-2). For examples of commitment to this requirement, see Hurley (1989: 146); Regan (1980: 124). We can apply this requirement to intrapersonal cooperation provided we understand 'group' such that a group could consist of several selves.

²⁹ See Woodard (2007: 90-5).

accidents when they drive abroad. However, it would be crazy for any individual UK driver to drive unilaterally on the right. Why? Because his doing so, together with others' predictable responses, would risk disastrous consequences. Call this the worry about recklessness.

We should not ignore facts about how others would respond to our actions. So, if we believe that facts about how they could respond help determine reasons, we'd better find some way of caring about both kinds of fact simultaneously. Whatever else Jim does, he'd better care about how Pedro would respond to each of his options. If he also cares about how Pedro could best respond, that had better not entirely displace the concern with how Pedro would respond.

Due to the way that the debate between actualists and possibilists is framed, we might rush to the assumption that this complex concern is somehow incoherent. As I mentioned in section 2, that debate is framed as being about the logic of ought statements. That's unfortunate, because on ordinary assumptions ought statements exclude each other. If Jim ought to shoot, it can't be that he also ought to not shoot.³⁰ However, we've been talking about reasons, not obligations. And on ordinary assumptions, reason statements do not exclude each other in the same way. If Jim has a reason to shoot, it could also be that he has a reason to not shoot. Reasons can conflict whereas obligations (we usually assume) cannot.

If we think of the issue between actualism and possibilism as usually framed, then, we may assume that we have to decide between caring about what others would do (actualism), or instead caring about what others could do (possibilism). But there's no need to choose one of these alternatives. We could care about both.

This would be a kind of pluralism, but not about values. It is pluralism instead about what we might call the *unit of action*: the stretch of action whose goodness or rightness generates a reason for action.³¹ According to pluralism of

³⁰ Some deny this. They claim that obligations can conflict, so that in some circumstances there is no way of avoiding acting wrongly. For example, see Nagel (1979: Chapter 5). However, since this is not how we usually think of obligations, framing the issue as one about obligations may well obscure the possibility of caring both about how others would respond and about how they could respond.

³¹ I am indebted to Susan Hurley's discussion of what she calls the 'unit of agency issue' (1989: 140–48). I have modified her terminology because on my view one can have pattern-based

this sort, the goodness or rightness of individual actions can generate reasons to perform them, but so too can the goodness or rightness of larger patterns. Pluralists believe that the causal properties *and* the parthood properties of individual actions ground reasons.³²

This is a way of combining concern with how agents would react to the actor's options with concern with how they could react to those options. For example, it is a way of combining concern with how Pedro would react to Jim's refusal to shoot with concern with how Pedro could react to that refusal. There's no need to choose between these concerns, provided we descend to the level of reasons instead of talking about obligations.

Pluralism of this sort is quite distinct from value pluralism, of course. Pluralism about the unit of action is a matter of how valued outcomes are associated with reasons for action; it isn't a matter of what has value. However, like pluralism about values this sort of pluralism can generate conflicting reasons. In Jim's case, this is just what we want. We set out to find an ethical view that could explain his strong reason to shoot as well as his strong countervailing reason not to shoot. By embracing pluralism about the unit of action we can do just that. His reason to shoot flows from the causal relationship between his options and the outcomes that would result. This causal relationship flows through Pedro's response, of course. In this respect, Pedro's significance is like the significance of a causally similar trolley, were that in Jim's environment. But pluralists can add that Jim also has a reason not to shoot, which flows from the parthood relationship between Jim's options and the patterns performable by him and Pedro taken together. These patterns include some in which Pedro responds to Jim in ways he could but would not. In this respect, Pedro's significance is quite unlike that of a causally similar trolley.

In our earlier case, we can say that you have a reason not to commit to help, which reflects the facts about how your future self would respond (by failing to help next weekend), but also a reason to commit, since that is part of the best pattern performable by your present and future selves.

reasons without extended forms of agency being either in existence or in prospect. See Woodard (2007: Chapter 4).

³² They might believe, also, that the intrinsic properties of actions can generate reasons. That idea has not been at issue here.

Note that pluralism about the unit of action can explain conflicts between reasons even where only one good is at stake. This is an attractive implication of this sort of view. Utilitarians, for example, could explain regret in this way, in at least some cases. Moreover, it is appealing to explain Jim's conflicting reasons as both stemming from the same value: the value of the innocent persons' lives. His reason to shoot and his reason not to shoot seem best captured in this way, since their loss dominates this scene. In contrast, views which have to appeal to distinct considerations, such as the value of Jim's own integrity, or which have to index judgements of value in some way, seem to get this case wrong. It is a distortion to say that Jim's reasons have these different sources. They both have the same source—in the one thing that matters in this case, which is the welfare of the hostages.

This illustrates a more general point. Often there are several ways we could try to explain some complex set of convictions. To some extent, we can choose where to distribute the complexity in our explanations. A common view amongst consequentialists is that we should try to keep all necessary complexity within the theory of value, so that we can keep the structural elements of consequentialism—those elements that generate claims about reasons or rightness when added to the theory of value—simple.³³ But this may sometimes result in implausible claims about value, as in Jim's case. Ideally, I take it, we would like an ethical theory with the following features: its structural components are plausible in themselves; its value claims are plausible in themselves; and its practical implications are plausible in themselves. Though Act Consequentialism's simple structure is plausible in itself, we may not be able to maximise overall plausibility by sticking with it. We should be open-minded about whether adding complexity in structure is preferable to adding complexity in value theory. In particular, we should not fetishise simple structure, if that leads us to make implausible claims about value.

Pluralism of the sort I've described is an attractive structure for cases like Jim's, in which we tend to have the conviction that there are conflicting reasons due to the presence in the actor's environment of some recalcitrant agent. This

³³ This may involve distinguishing more finely between types of actions, or in some other way adding complexity to the theory of value. For careful defence of this general strategy, see Portmore (2007). I am grateful to Guy Fletcher and Douglas Portmore for discussion here.

recalcitrant agent may be one's own future self, as in the cases discussed in the debate between actualists and possibilists, or it may be a different individual, as with Jim and Pedro. In such cases we seem to have a dual concern with the recalcitrant agent's response: how he would respond matters, but how he could respond seems to matter too. I've explained how both these concerns could operate together, reflecting the significance both of the causal relations between individual actions and valued outcomes, and of the parthood relations between individual actions and favoured patterns of action.

5. Conclusion

If we wish to explain both of Jim's reasons, we should take pluralism about the unit of action very seriously. Jim's reason to shoot plausibly reflects the importance of what Pedro would do, while Jim's reason not to shoot plausibly reflects the importance of what Pedro could do. This sort of explanation seems to have at least as much promise as the main competitors, which focus on what Jim does.

There are hard problems facing the sort of pluralism I've described. To get a full-fledged theory, one would need to specify what makes patterns eligible to support reasons in a way that is both plausible and avoids endorsing the Willingness Requirement. To be plausible, this account should answer the worry about mere agglomerations, and enable us to answer the worry about recklessness.³⁴ Supposing we could do all that, we'd also want to know something about how the reasons associated with different units of action interact with each other—just as we'd like to know how the reasons associated with different values interact with each other.

One of the advantages of this sort of view is that it explains why a causally similar trolley would not have served Williams's purposes just as well. Were Jim faced with a trolley that would kill none if he shoots one, or kill twenty if he refuses to shoot, we would find it less troubling to say that he should shoot. His reasons in this case seem different to those in the case involving Pedro. However,

³⁴ I attempt these tasks elsewhere (Woodard 2007: Chapters 5-6).

the way the trolley would respond to each of Jim's choices is, in relevant respects, just the same as the way that Pedro would respond. Moreover, any special relationship that exists between Jim and his own actions or point of view is constant across these two cases. So, views that focus on what Jim does seem ill equipped to explain our different reactions to the two cases.

A critic might object as follows. The fact that Pedro is an agent surfaces within the response he would make to Jim's choices. We do not need to look any further, or to consider what Pedro could do. The important difference between Pedro and the trolley is that, if Jim refuses to shoot the trolley would *kill* twenty, while in the other case Pedro would *murder* twenty. Pedro's agency triggers the application of different moral concepts, which enables us to distinguish between the cases even if we are concerned only with how Pedro would respond, and not at all with how he could respond.

That is true. However, consider a final case. Suppose Jim faces *Frankfurt Pedro*. If Jim shoots, Frankfurt Pedro will shoot none. If Jim refuses, Frankfurt Pedro will shoot twenty. Not only that, but since he is a responsible agent, Frankfurt Pedro would murder twenty. However, he could not respond in any other way to Jim's refusal—because a counterfactual intervention mechanism would prevent him if he tried to do so.³⁵

If the critic is right, we should have the same convictions about Jim's choice when faced with Frankfurt Pedro as we do about Jim's choice when faced with Ordinary Pedro. In both cases, Pedro would not merely kill twenty were Jim to refuse to shoot; he would murder them. But Jim's reasons seem responsive to the fact that Frankfurt Pedro could not do otherwise. While Frankfurt Pedro is certainly not exactly like a runaway trolley, he is more similar to such a trolley than is Ordinary Pedro. If, as I suppose, we think that Jim's reasons reflect the difference between these cases in any degree, that is evidence that we care not only about what Pedro would do, but also about what he could do.

³⁵ See Frankfurt (1969) and Hurley (2003: 16–20 and Chapter 2). Of course there is controversy about whether Frankfurt-type cases really show that responsibility does not require that the agent concerned could have done otherwise. If they do not it is harder for me to respond to the objection at hand. But even so, we should be able to distinguish between concern with the contrast between murder and other kinds of killing, on one hand, and concern with the contrast between how other agents would respond and how they could respond, on the other.

References

- Bennett, J. (1995). *The Act Itself*. Oxford: Clarendon Press.
- Bergström, L. (1966). *The Alternatives and Consequences of Actions*. Stockholm: Almqvist and Wiskell.
- Dancy, J. (2004). *Ethics Without Principles*. Oxford: Oxford University Press.
- Frankfurt, H. G. (1969). 'Alternate Possibilities and Moral Responsibility', *Journal of Philosophy*, 66: 829–39.
- Hollis, M. (1983). 'Jim and the Indians', *Analysis*, 43: 36-9.
- Hurley, S. L. (1989). *Natural Reasons*. New York: Oxford University Press.
- Hurley, S. L. (2003). *Justice, Luck, and Knowledge*. Cambridge MA: Harvard University Press.
- Jackson, F. (1987). 'Group Morality', in Pettit, P., Sylvan, R. and Norman, J. (eds) *Metaphysics and Morality. Essays in Honour of J. J. C. Smart*. Oxford: Basil Blackwell, 91–110.
- Jackson, F. and Pargetter, R. (1986). 'Oughts, Options, and Actualism', *The Philosophical Review*, 95: 233–55.
- Kagan, S. (1989). *The Limits of Morality*. Oxford: Clarendon Press.
- Kamm, F. M. (1992). 'Non-Consequentialism, the Person as an End-in-Itself, and the Significance of Status', *Philosophy and Public Affairs*, 21: 354-89.
- Kamm, F. M. (2006). *Intricate Ethics: Rights, Responsibilities, and Permissible Harm*. Cary, NC: Oxford University Press.
- Mulgan, T. (2001). *The Demands of Consequentialism*. Oxford: Clarendon Press.
- Nagel, T. (1979). *Mortal Questions*. Cambridge: Cambridge University Press.
- Parfit, D. (1987). *Reasons and Persons*, reprint with corrections. Oxford: Clarendon Press.
- Parfit, D. (2007). *Climbing the Mountain*, unpublished manuscript, accessed at http://individual.utoronto.ca/stafforini/parfit/parfit-climbing_the_mountain.pdf on 21.03.07.
- Portmore, D. W. (2007). 'Consequentializing Moral Theories', *Pacific Philosophical Quarterly*, 88: 39-73.
- Quinn, W. (1989). 'Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing', *The Philosophical Review*, 98: 287-312.

- Regan, D. (1980). *Utilitarianism and Co-operation*. Oxford: Clarendon Press.
- Scanlon, T. M. (1998). *What We Owe to Each Other*. Cambridge MA: Harvard University Press.
- Scheffler, S. (1982). *The Rejection of Consequentialism*. Oxford: Clarendon Press.
- Scheffler, S. (2004). 'Doing and Allowing', *Ethics*, 114: 215-239.
- Williams, B. (1973). 'A Critique of Utilitarianism', in Smart, J. J. C. and Williams, B., *Utilitarianism: for and against*. Cambridge: Cambridge University Press, 75–150.
- Williams, B. (1995). 'Acts and Omissions, Doing and Not Doing', in Hursthouse, R., Lawrence, G. and Quinn, W. (eds) *Virtues and Reasons. Philippa Foot and Moral Theory*. Oxford: Clarendon Press, 331–40.
- Woodard, C. (2003). 'Group-Based Reasons for Action', *Ethical Theory and Moral Practice*, 6: 215–29.
- Woodard, C. (2007). *Reasons, Patterns, and Cooperation*. New York: Routledge.
- Woodard, C. (2008). 'A New Argument Against Rule Consequentialism', *Ethical Theory and Moral Practice*, forthcoming.
- Zimmerman, M. J. (1996). *The Concept of Moral Obligation*. Cambridge: Cambridge University Press.