The Second Mistake in Moral Mathematics is Not About the Worth of Mere Participation

Petersson, Björn

Published in:
Utilitas

DOI:
10.1017/S0953820804001189

2004

The Second Mistake in Moral Mathematics is not about the Worth of Mere Participation

BJÖRN PETERSSON
Lund University

'The Second Mistake' (TSM) is to think that if an act is right or wrong because of its effects, the only relevant effects are the effects of this particular act. This is not (as some think) a truism, since 'the effects of this particular act' and 'its effects' need not co-refer. Derek Parfit's rejection of TSM is based mainly on intuitions concerning sets of acts that over-determine certain harms. In these cases, each act belongs to the relevant set in virtue of a causal relation (other than marginal contribution) to a specific harmful event. This feature may make an act wrong, in a fashion consequentialists could admit. That explication of TSM does not rely on the questionable assumption that the *set* of acts is what harms here. Independently of this, there are several other reasons to prefer it to the 'mere participation' approach.

I. INTRODUCTION

'(The Second Mistake) If some act is right or wrong *because of its effects*, the only relevant effects are the effects of this particular act'.[1] Derek Parfit claims that this sentence expresses a tempting but mistaken way of reasoning about morally relevant consequences. In real life, people often commit this mistake, he says, e.g. when they consider things like soil-erosion, over-fishing, inflation, depletion, over-farming etc.[2] The Second Mistake (TSM) is wrong because it overlooks some harmful or benefiting effects that we produce together with other people, according to Parfit.

Frank Jackson argues that the 'mistake' in its original formulation is a truism, and Ben Eggleston agrees, 'for what could the phrase "*its effects*" refer to, if not the effects of the particular act'.[3] They take it for granted that the supposedly trivializing 'its' is simply 'a slip'.[4] In that case, Parfit repeats the slip in several places, e.g. when he discusses TSM at the end of the book,[5] in his comments to Gruzalski,[6] and when he expresses what he considers to be the correct view. An action that

[1] Derek Parfit, *Reasons and Persons* (Oxford, 1984), p. 70.
[2] Ibid., p. 444.
[3] Ben Eggleston, 'Should Consequentialists Make Parfit's Second Mistake? A Refutation of Jackson', *Australasian Journal of Philosophy* 78 (2000), p. 2.
[4] Frank Jackson, 'Which Effects', *Reading Parfit*, ed. J. Dancy (Oxford, 1997), p. 52.
[5] Parfit, *Reasons and Persons*, p. 443.
[6] Derek Parfit, 'Comments', *Ethics* 96 (1996), p. 849.

DOI: 10.1017/S0953820804001189
Utilitas Vol. 16, No. 3, November 2004
Printed in the United Kingdom

does not harm other people may, he says, still be wrong because of *its* harmful effects on other people.[7]

Jackson argues that even when the error is corrected, TSM is no mistake that consequentialists should avoid. Eggleston defends the rejection of TSM, although he thinks that Parfit's rejection of it commits him to accepting that mere participation can matter morally. This, Eggleston claims, is inconsistent with Parfit's reasons for rejecting 'the Share-of-the-Total View'.[8]

In this article, I advocate an explication of TSM that is consistent with Parfit's original formulation without being trivially true. 'Its effects' and 'the effects of this particular act' need not co-refer. My main purpose is to show that we can reject TSM and admit that an act can be wrong when it is one in a set of acts that together harm people, without assigning worth to mere participation. I am not making any exegetical claims about which reasons that Parfit actually had in mind when he rejected TSM. Nor is my point that TSM really is a mistake. What I want to show is that the commitments that might come with rejecting TSM are irrelevant to the worth of mere participation.

The assumption that mere participation can matter is usually contrasted with the view that individual marginal contribution is the only thing that matters (to the assessment of the moral worth of an action). But these alternatives are not exhaustive. When some harmful event attracts our attention and we find that it is produced by intentional actions, we tend to condemn the actions that marginally contribute to this harm, i.e. the acts that are such that had they not been performed, less of that harm would have occurred. There are cases in which there is no individual action of which this is true, although we can trace the cause of the harm to a certain set of acts. To assume that membership in such a set can make an individual action wrong is to admit that individual marginal contribution is not the only thing that matters. This, I will argue, does not imply that mere participation can matter, i.e. that an action can be morally tainted by an effect of a group's actions, regardless of the causal relations between this individual act and the effect in question.

When people argue that mere participation can matter to the rightness or wrongness of an act, the idea is that even though the act makes no marginal contribution to a morally compromising effect of a set of acts, the mere fact that the act is one in the set can affect its moral worth. This way of reasoning presupposes that participation is distinct

[7] Parfit, *Reasons and Persons*, p. 82.

[8] Eggleston defends the rejection of TSM in 'Should Consequentialists Make Parfit's Second Mistake?' and discusses the possible inconsistency in 'Does Participation Matter? An Inconsistency in Parfit's Moral Mathematics', *Utilitas* 15 (2003).

from marginal contribution. Other features than degree of marginal contribution must determine degree of membership in the relevant set. What are these features? I see two main possibilities. Either, an act is a member of the relevant set solely in virtue of relations to other acts in the set. Or, the act is a member because it shares with the other acts some causal relation to the harmful/benefiting effect in question, other than marginal contribution to it.

Parfit's reason for rejecting the TSM is that he believes that an act can be wrong because it is one in a set of acts that has harmful or benefiting effects, even though 'the effects of the particular act' are morally neutral. Eggleston argues that this way of thinking commits Parfit to the view that mere participation can matter. Since Eggleston regards the original formulation of TSM as a truism, he seems to overlook the possibility that an agent's participation could be a matter of the act's effects. This is because, if membership in the relevant set of acts was determined by effects of the individual act (other than marginal contribution to the overall effect), there would be a sense in which 'its effects' could be morally relevant, although 'the effects of this particular act' (understood in terms of marginal contribution) are morally neutral.

So, we should distinguish between three views. The first is that an act's marginal contribution to harmful/benefiting effects is the only thing that can matter. The second idea is that other causal relations between the act and the set's overall harmful/benefiting effects may affect the act's rightness or wrongness. The third possibility is that relations between the act and other acts in a set of acts that has harmful/benefiting effects can matter to its rightness or wrongness, regardless of the causal relations between the act and the harmful/benefiting effects of the set.

I will suppose that 'mere participation can matter' is an expression of the third of these views. When Eggleston characterizes TSM as a mistake of overlooking the worth of mere participation, because it is 'to consider an act in isolation from other acts with which it is connected', it seems clear that it is the third view he has in mind.[9] I will argue that although the thought experiments Parfit appeals to when he rejects the TSM give us reasons to doubt that marginal contribution is the only thing that matters, they do not support the view that mere participation can matter.

Two thought experiments are essential to Parfit's rejection of TSM, and I will mainly confine my attention to this part of Parfit's argument. *Case One* describes an over-determined effect – two people kill me

[9] Eggleston, 'Does Participation Matter?', p. 95.

simultaneously and each act is sufficient for my death at a certain time. In *Case Two*, one act's causing a certain effect is pre-empted by another act's effect – X performs an act that will kill me and Y thereafter kills me just before X's act has its full effect.

Section II in this article is an attempt to eliminate some confusion that might arise concerning the causal relations described in these examples. I suggest that one form of mistake could stem from a common but in some cases too simple counterfactual assumption about causal relations. With the aid of Lewis's notion of fragile versions of events, I also make clear how an act can be among the causes of an effect of a set of acts to which it belongs, without marginally contributing to the occurrence of this effect.

In section III, I question Parfit's assumption that what makes the individual act wrong in these cases is that it belongs to a set of acts that harms someone. In Parfit's view, 'an act harms someone if its consequence is that someone is harmed more'.[10] 'More', I take it, implicitly means 'more than if the act had not been performed'. When discussing cases of the type mentioned, Parfit also assumes that a set of acts harms someone if it is the smallest set such that if none of the acts had been performed, this person would not have been harmed.[11] Membership in such a set can make an act wrong, he thinks, even if the individual act harms no one. I argue that on the criteria of 'harming someone' that Parfit applies to individual acts, sets of the mentioned kind cannot be said to harm someone. So, if it is wrong to participate in such a set, it is not because the set of acts harms someone.

Section IV gives a positive explanation of why one nevertheless may consider it wrong to perform an act that belongs to a set of the mentioned kind. It proceeds by focusing criteria of membership in the relevant kind of sets. My contention is that in cases of the sort described, we pick out members in terms of their tendency to cause the harmful effect that makes us morally concerned in the first place. This requires a specific non-frequential notion of a single case causal tendency. I argue that it is quite compatible with act-consequentialism to regard acts as wrong owing to their tendency to cause harm in this special sense. But then, what makes it wrong to perform an act that belongs to 'the smallest set of acts such that if none of the acts had been performed, this person would not have been harmed' is not the mere participation in that set, but those features of the act that make it qualified for membership. Furthermore, facts about those features are about the effects of the act. This means that there is a sense in which an act may be wrong because of its effects, although it makes no marginal

[10] Parfit, *Reasons and Persons*, p. 69.
[11] Ibid., p. 71.

contribution to the occurrence of any harmful event. Interpreted in that way, 'its effects' and 'the effects of the particular act' may come apart. Those who reject TSM, because of thought experiments of the kind Parfit appeals to, are committed to some view of this kind, rather than to the idea that mere participation matters.

Section V adds another reason for preferring my suggested reading of Parfit's principles. It explains how we can delimit an otherwise random set of wrong acts solely in terms of their connection to a certain harmful effect (pollution, extinction of fish supply, etc.) while thinking that each specific action in a sense is harmless. In these cases, what makes an action a member of the set is that it is somehow causally connected to the production of a certain harmful event. Its role in that causal process is what makes us pick it out as a member. Unless we have a way of stating about the act that its effects are what make it wrong in such cases, it is hard to see how we could claim that that agent is participating at all – that his action is a member of the relevant set of acts.

Section VI is an attempt to illustrate that TSM is a special case of a more general principle. This more general mistake is not exclusively about sets of actions performed by groups of agents. We can make the same kind of mistake when we attempt to evaluate whether other kinds of events are good or bad for us. In the latter case, it is even more evident that the mistake is not about the worth of participation.

Parfit's thought experiments on 'imperceptible harms' are discussed in section VII. Although these examples may support the rejection of TSM, they do not support the 'mere participation' view.

Section VIII briefly discusses Jackson's defence of TSM, and in section IX it is argued that Parfit's rejection of 'the Share-of-the-Total View' is compatible with the view that mere participation can matter. We can reject the Share-of-the-Total View as well as TSM, and still leave it open whether mere participation matters.

Conclusion: The question of whether mere participation can matter morally must be settled on other grounds than those Parfit appeals to in these two rebuttals.

II. WHO IS CAUSALLY RESPONSIBLE FOR MY DEATH?

A phrase like 'C caused E' would in daily speech probably express the thought that C alone caused E, or at least that C was an important cause of E, perhaps the 'triggering' cause among a host of necessary causal background conditions. In order to avoid such implicit assumptions I will instead discuss criteria for assuming that 'C is causally responsible for E'. To be causally responsible for an event is simply to be a part of its causal genesis. If you accidentally stumble,

fall, and break my vase, you are causally responsible for its breaking in the same weak and purely descriptive sense in which (e.g.) the acid may be said to be causally responsible for making the metal corrode.

'C is causally responsible for E' does not imply that C alone is causally responsible, or that C is responsible in some especially important way. If ten people each perform an act that is necessary to produce a certain result, and their acts together are sufficient for it, then each is causally responsible for that result.

One mistake illuminated by Parfit's thought experiments stems from equating causal responsibility with marginal contribution. Take *Case Two*, which is a case of pre-emption. X performs an act that is sufficient to kill me. Y thereafter kills me, say, by shooting me. If she hadn't fired her gun, I would have died a few minutes later from the effects of X's act. Suppose that Y with certainty foresees the events that will follow after her firing the gun until I am dead, as well as the events that would occur if she hadn't. Now, she wants to know whether it really will be her action that causes my death this evening.

As Parfit's discussion indicates, she might be tempted to think that she has no causal responsibility for my death. On what grounds? What she seems to be interested in is her marginal contribution to my dying, i.e. the difference her action will make to the occurrence of that event. And the most natural way of thinking that she might apply then is to reason in counterfactual terms. This way of thinking appears to be what lies behind one conclusion Parfit thinks one might be tempted towards in such situations – that if an effect would occur in the absence of my action, then my action cannot make me causally responsible for that effect. In the simplest form: 'C is causally responsible for E' implies 'if C does not occur, E will not happen'. If we allow for sufficient slack in our criteria of identity of the event 'my dying this evening', she may then conclude that her shooting will not make her causally responsible for that event.

This way of thinking is not apparently flawed or absurd when applied to more straightforward causal processes. In daily life, we often reason like this, and as it happens, it corresponds to an influential tradition in the philosophical analysis of causation. As many have stressed, though, the counterfactual account of causality must be made more refined than above, precisely because of difficulties in handling cases like the one described.[12] I am not here claiming that the counterfactual analysis is the correct analysis of causality. My suggestion is merely that the most reasonable explanation of why a person like Y might be tempted to think that she has no causal responsibility for my dying is that she

[12] See David Lewis, 'Causation as Influence', *Journal of Philosophy* 97 (2000).

makes a counterfactual assumption as described, and that this is a common way of thinking about causal responsibility in daily life.

If we hold on to the counterfactual criterion, how could Y conclude that she is causally responsible for my death? One line of reasoning, elaborated e.g. by Lewis[13] and McDermott,[14] makes this conclusion compatible with the counterfactual criterion by focusing on criteria of event-identity. We could say that without Y's gunfire, I would still have died that evening, but that this would have been a different dying. In Lewis's terminology, we would then assume 'my dying' to be a fragile event. I would die in almost the same way and at almost the same time without Y's shot, but the slight delay in time and perhaps some small change, e.g. in tissue damage, would suffice to make that dying another event.

As Lewis stresses, there are no firm conceptual criteria in ordinary language when it comes to the identity of events.

> How much delay or change do we think it takes to replace an event by an altogether different event, and not just by a different version of the same event? An urgent question, if we want to analyze causation in terms of the dependence of whether one event occurs on whether another event occurs. Yet once we attend to the question, we surely see that it has no determinate answer. We have not made up our minds; and if we do presuppose sometimes one answer and sometimes another, we are entirely within our linguistic rights.[15]

We could stipulate that any sort of change, hastening, or delay, however small, will turn one event into another. Maximal fragility would exclude the possibility of an event coming in more than one version. It would always be the case about any event, that unless it had been caused exactly at the time and in the way it was, it would not have occurred.

In many contexts it would be linguistically inappropriate to make events so fragile as to exclude versions. If we discuss, say, what caused the extinction of the dinosaurs, or the Second World War, surely we can think of a variety of different versions of those very events, with slightly differing possible time-spans and courses of causation. You and I may disagree about how and when they took place, exactly, while we do not doubt that we refer to the same events. Intuitively, it also seems that we quite often might regard 'C caused E' as true, while we still doubt that 'If C had not occurred, E would not have occurred' is true.

As Lewis makes clear (developing an amendment of the counterfactual analysis, suggested by G. E. Paul), the counterfactual account can still accommodate cases like the one described by

[13] Ibid.
[14] M. McDermott, 'Causation: Influence versus Sufficiency', *Journal of Philosophy* 99 (2002).
[15] Lewis, 'Causation as Influence', p. 186.

employing the notion of fragility. We should just 'relocate the fragility not in the event itself but rather in a tailor-made proposition about that event, that will be a proposition about how and when and whether the effect occurs'.[16] We should not understand 'C will cause E' as implying that 'if C does not occur, E will not happen', but rather 'if C does not occur, E will not happen, or will not happen in the same way or at the same time'. Y's causing my death this evening then not only depends on whether this event would occur without her gunfire, but on whether it would occur at the same time and in the same manner without it.

In that way, Y should still say, as Parfit wants her to, that what she foresees in *Case Two* is that she is causally responsible for my death, and that X is not. Unless she fires, the event 'my dying this evening' will nevertheless occur, but not in the same manner or at the same time as if she fires. She causes my death, but she does not marginally contribute to it. Marginal contribution to an event is a matter of making a difference to the *occurrence* of that event.

One difficulty with making versions of events maximally fragile is that it seems to 'open the gate to a flood of spurious causes'.[17] A cool breeze that evening may have affected the temperature and course of Y's bullet ever so slightly. If the absence of that minute force would result in a different version of 'my dying this evening', we would have to say that the breeze was causally responsible for that event. If we do not want to admit that, we will have to allow for some slack in our criteria of the identity of versions as well. In some unusual contexts, it might be right, as Lewis says, to attend to such small differences and regard all the minute influences as joint causes of the effect. In most other contexts, we simply regard them as negligible.

Let me, now, turn to a case of over-determination. This is a slightly more specified version of *Case One*, the thought experiment Parfit initially employs to establish that TSM is a mistake.[18]

> X and Y simultaneously shoot me. Either shot would, on its own, have been sufficient to kill me in the same manner and at the same time.

If this description is taken literally, it means that the same version of the same event would occur if Y fired her gun, if X fired his gun, or if both fired their guns. But to begin with, 'in the same manner' cannot mean 'in exactly the same manner, down to the most minute detail'. In this case there will necessarily be some small differences between the effects in the three cases. Even if we assume that in all three cases

[16] Ibid., p. 187.
[17] Ibid., p. 188.
[18] Parfit, *Reasons and Persons*, p. 70.

a piece of lead of a certain size, density and speed will enter my body from a certain angle and result in some specific bodily damage at a certain point in time, the numerical identity of the lead that enters my body will depend on whether it stems from Y's, X's, or both guns, for instance.

Either, we could still take the case description literally, and assume that these differences, like the difference the breeze would make, are below the level of negligibility. Which specific lump of lead enters my body then does not make a difference to how and when I die. This would force us to conclude that it is indeterminate whether I will be killed by X's or by Y's action, or by some combination of both. This indeterminacy is not due to some metaphysical constraint; nor is it merely an epistemic limitation. It is conceptual and follows from our level of specification when it comes to describing 'how' and 'when'.

Or, we could regard this as one of the unusual cases in which even the most minute change is relevant, and attempt to compare the sizes of the minimal differences that Y's abstention or X's abstention would make to the actual case. Why shouldn't we? After all, in this thought experiment, we may assume that we know those details, and that Y and X are able to predict them before they shoot. If Y knows that the difference her abstention would make to the actual case is exactly as small as the difference X's abstention would make, she should conclude that they jointly will cause that event, and that each will have causal responsibility for it.

Either way, it would be fallacious of X or Y to conclude that his/her shooting me will not make him/her causally responsible for my death, from the correct assumption that the event caused will occur anyway. Each could, though, draw the conclusion that it is indeterminate whether his/her shot will make him/her causally responsible. Alternatively, they could infer that each one has causal responsibility for the event.

The description above does not commit me to the view that genuine overdetermination is possible, i.e. that a single event can have more than one sufficient operative cause. But suppose this is possible. Let me modify the example somewhat to illuminate that possibility. Suppose X and Y simultaneously fire their guns towards a mechanism that drops a heavy weight on me, and that this weight will cause my death in exactly the same manner, down to the most minute detail, independently of whether X's, Y's or both guns are fired. Assume also that we are allowed to limit our description of the relevant event so that the details of how the mechanism is affected lie outside the event. If each bullet is sufficient and operative in releasing the weight, then each should conclude that his/her act causes the event. To assume that there is genuine overdetermination is to deny that the fact that X's act causes

E excludes the possibility that Y's act causes E. So, the conclusion that each is causally responsible in the overdetermination case could be correct independently of whether there is genuine overdetermination or not.

In *Case Two*, Parfit adds more salient changes to what would have happened if Y had not shot me. My death would then be delayed by a few minutes, it would be more painful, and there would be considerable differences in bodily damage. With some generosity, we might admit that these changes are insufficient to turn the relevant event into another, but it would be absurd to claim that they make no difference to how and when this event occurs. So, although the event would occur in the absence of Y's act, it is Y alone who is causally responsible for my death in this case.

The upshot of this section is that there is one straightforward way in which an act can be causally related to an effect of a set of acts to which it belongs, without marginally contributing to this effect: It can be one of its causes. In the pre-emption case, Y alone is causally responsible for my death without having marginally contributed to it. In the overdetermination case, each is causally responsible for my death (or it is indeterminate which one is responsible) though neither makes any marginal contribution to it.

III. WHO HARMS ME?

(C6) An act benefits someone if its consequence is that someone is benefited more. An act harms someone if its consequence is that someone is harmed more. The act that benefits people most is the act whose consequence is that people are benefited most.[19]

(C7) Even if an act harms no one, this act may be wrong because it is one of a *set* of acts that *together* harm other people. Similarly, even if some act benefits no one, it can be what someone ought to do, because it is one of a set of acts that together benefit other people.[20]

(C8) When some group together harm or benefit other people, this group is the *smallest* group of whom it is true that, if they had all acted differently, the other people would not have been harmed, or benefited.[21]

According to C6, the question of whether an act harms/benefits is not simply a matter of whether the action is causally responsible for a

[19] Ibid., p. 69.
[20] Ibid., p. 70.
[21] Ibid., p. 71–2.

certain harmful/benefiting event, nor is it merely a question of whether the act marginally contributes to the occurrence of some specific harmful/benefiting event. C6 requires us to compare the amount of harm/benefit caused with the amount of harm/benefit that would have been produced without that act.

I assume that it is implicit in C6 that the relevant counterfactual alternative to compare when we consider whether someone is harmed more concerns the amount of harm that would have occurred without that act, other things being equal. To find out whether an act harms/benefits most, we have to compare the amount of harm that would have occurred on any of the available alternative actions.

The verdict of C6 is that neither X nor Y harms me in *Case One* and *Case Two*. But since I am harmed, who harms me? Parfit's suggestion is that X and Y harm me together, i.e. that the set consisting of these two acts harm me. The set of acts that harms me, says Parfit, is the smallest set of acts such that if none of the acts had been performed, I would not have been harmed. His point is that if we do not commit TSM, we will realize that an action can be wrong because it belongs to such a set.

However, if we apply C6 to the set of acts in this case, it does not follow that this set harms me. Parfit seems to presuppose that if the set of acts had not been performed, then none of its members would have occurred. In that case, if the set had not occurred, I would not have been harmed. But that assumption is, at best, arbitrary. The actual case is that both shoot me. The relevant counterfactual case to compare when we discuss the harmfulness of X's act is supposed to be a case in which Y shoots me (rather than a case in which neither shoots me). That is the case which comes closest to the actual case, absent X's act. But why, then, should the relevant counterfactual alternative to compare the set with be a case where neither shoots me? The 'other things equal' clause requires us to compare the counterfactual alternative that comes closest to the actual one, absent the set of acts. Would it not be more reasonable, then, to assume that if the set of acts had not occurred, then all but one of its members would have occurred? At least, it is not clear why this possibility should be excluded. This would mean that neither X's act, nor Y's act, nor the set of X's and Y's acts, harms me.[22]

The fact that a set is the smallest set of acts such that if none of the acts had been performed, I would not have been harmed, does not by itself imply that the set harms me on the C6 criterion of what it is to harm someone. If we assume that 'the smallest set of acts such that if none of the acts had been performed, this person would not have been

[22] A comment by Wlodek Rabinowicz brought these doubts (about the set's harming) to my mind.

harmed' would not occur, it does not follow that this person would not be harmed. In this counterfactual assumption, it would be arbitrary to exclude the possibility that some of the set's former members occur when the set does not. Therefore, 'the smallest set of acts such that . . .' does not necessarily harm on the criterion of harming that Parfit applies to acts. The verdict of C6 should be that neither X's act nor Y's act, nor the set of these two acts, harms me.

Who or what harms me then? I doubt that this question is appropriate if we hold on to the comparative criterion of harming someone that is stated in C6. We should simply admit that there are cases in which this question has no definite answer. These are cases such that we cannot, in principle, say about the set or any of its single members, that if it had not occurred I would not have been harmed, although we know that the cause of some harmful event that strikes me can be found in that set.

(There are other common ways of talking about harming that might explain why the question sounds meaningful even in cases like those discussed. When asking who or what harms me, we may (e.g.) want to know what causes the harmful event in question. In that sense of harming, the question is less obscure. The answer would then be that in *Case One*, either each act causes that event, or it is indeterminate which one causes it. In *Case Two*, Y's act causes the harmful event and X's does not. But that is not what harming means according to the comparative criterion in C6.)

Again, on the criteria of 'harming someone' that Parfit applies to individual acts, it does not follow that a set of the mentioned kind harms someone. So, if it is wrong to participate in such a set, it is not because the set of acts harms someone.

Before closing this section I will briefly consider another possible solution. The disjunctive state of affairs 'X kills me or Y kills me' seems to harm me in the sense considered here.[23] If this state does not obtain, I will not be harmed. Could this be the sort of answer required to explain why it is wrong to act as X or Y does?

Let us assume with Chisholm and others that states of affairs resemble propositions and that it is unproblematic to assign logical connectives to them.[24] Assume also that there is a plausible sense in which disjunctive states of affairs can be said to occur or obtain. Then it seems reasonable to say that it is wrong to promote the occurrence

[23] I owe this observation to *Utilitas*'s referee.

[24] See e.g. R. M. Chisholm, 'The Intrinsic Value in Disjunctive States of Affairs', *Nous* 9 (1975) p. 295; E. Carlson, 'The Intrinsic Value of Non-basic States of Affairs', *Philosophical Studies* 85 (1997), p. 95; and M. J. Zimmerman, 'Evaluatively Incomplete States of Affairs', *Philosophical Studies* 43 (1983), p. 211.

of a disjunctive state that harms someone. Note that for a disjunction of acts to harm someone in C6's sense, it is required that someone is harmed more if any of the disjuncts is realized than if none of them is.

Who or what makes a disjunctive state of that special kind occur in the case under consideration? Since one of the disjuncts suffices to make the disjunctive state obtain, the occurrence of that state is overdetermined here. We cannot say about X's act, Y's act, or the set, that it is what makes the state obtain. All we can say is that any of these three possible occurrences is sufficient to make it so. The disjunctive state might occur in the absence of any of them. (If Parfit had been right about the set's harming me, there would be some plausibility in assuming that each act's contributing to the occurrence of the set might make the act wrong. But we have no similar way of relating the occurrence of the disjunctive state to each act.)

In other words, even if we admit that the state 'X or Y kills me' is what harms me, we need to know who or what made this state obtain in order to assign wrongness. This manoeuvre would just generate a new question that cannot have a definite answer in cases like these.

IV. WHY IS IT WRONG TO BELONG TO THE SET?

Suppose we agree with Parfit that it can be wrong to perform an act that belongs to the smallest set of acts such that if none of the acts had been performed, I would not have been harmed. As we have seen, we cannot argue that the act is wrong because it belongs to a set that harms me in C6's sense. Why, then, might such an act be wrong? Consider the facts of *Case One* again. My death is the event that is supposed to make us morally concerned in the first place. This event is such that had it not occurred, I would not have been harmed. If any of the two acts are performed, this event will occur. Each act's sufficiency for the event is what makes it a member of the relevant set. What makes us assume that a certain act is sufficient for that event is that we know that it will make a bullet enter my body and cause bodily damage of the kind that kills me. Each act initiates a causal tendency towards my death – a chain of events that predictably causes my death unless this event becomes caused by the other act or by both jointly. In that sense, the effects of each act are what make it a member of the set.

My suggestion is that in cases like those described, the reason that an act can be wrong when it is one in a set of the kind described is that it can be wrong to perform an act that tends to cause an event which harms in C6's sense. An event that harms me in that sense is an event such that if it does not happen, I am not harmed.

Perhaps this introduction of causal tendencies appears problematic. To begin with, it may seem to beg the question against

act-consequentialism. If 'tendency' is understood frequentially, the principle that it can be wrong to perform an act that tends to cause an event which harms in effect says that it may be wrong to perform acts of the type that usually cause harmful events. That would be a rule-consequentialist principle. It must be stressed, therefore, that 'tendency' here denotes a single-case tendency.

Consider J. S. Mill's famous proposal that 'actions are right in proportion as they tend to promote happiness, wrong as they tend to promote the reverse of happiness'. He could mean, like rule-utilitarians, that actions are right if they belong to a type of acting, which in 'normal' circumstances promotes happiness – permitting that individual actions are right in spite of their failure to promote happiness. Or he might claim that the rightness of an action is proportionate to the degree of happiness-promotion it tends to yield in every individual case.[25] It is the latter interpretation of tendency that I have in mind here.

A single case tendency is an occurrent activity and should be distinguished from a dispositional state. A brittle object, like the coffee-cup in front of me, has, through its existence, a disposition to break. It may also, now, tend to break. In ordinary speech, both these truths can be covered by 'The cup tends to break'. In a similar way, the statement 'I tend to drink too much coffee' is ambiguous: It might describe a passive state, identified in terms of frequency of coffee-drinking or hypothetical prediction of coffee-breaks given normal circumstances, or it could denote an occurrent activity. I might be disposed to drink too much coffee without, now, tending to do it. It can also be true that I, now, tend to drink too much coffee although I have no disposition to do it.

What a modern philosopher would call a 'dispositional' account of tendencies blurs this distinction between occurrence and state by failing to allow for the possibility that 'tending to do something' can equally well denote either an occurrence or a state.[26]

I said that each act in *Case One* initiates a causal tendency towards my death. 'Towards' implies direction, and some may also worry that the concept of a single case tendency employed here turns tendencies into occult powers. (This is a confusion T. S. Champlin attributes to Mill and Descartes.) However, I am merely claiming that statements about tendencies implicitly indicate a foreseen possible result, from the spectator's point of view. The result is implicitly understood as

[25] See T. S. Champlin and A. D. M. Walker, 'Tendencies, Frequencies, and Classical Utilitarianism', *Analysis* 35 (1974), and Björn Petersson, 'Causal Tendencies', in *Belief and Desire, The Standard Model of Intentional Action* (Lund, 2000), sec. 2.2.

[26] T. S. Champlin, 'Tendencies', *Proceedings of the Aristotelian Society* 91 (1991), p. 127.

an unmodified claim, which the tendency-statement is supposed to weaken. Although that may make the concept of a tendency teleological in some sense, it does not imply that there are any pre-imagined goals in the tendencies themselves.

The required notion is that of a triggered causal process, which would produce a certain result, if the possibility of shortfalls were not apparent. The (single-case non-statistical) tendency of the ice to break under the foolhardy skater has a direction in this profane sense. When we attribute that tendency to the ice, we have the result of the completed causal chain in mind. In other words, we can talk about the direction of a causal tendency without turning tendencies into agents or magic forces. This use of the term presupposes only that it is applied to a process, which can be thought of as producing a specific result, and that the process, but not necessarily the result, can be imagined as partly fulfilled.

Consider, again, the act-utilitarian interpretation of Mill's statement 'actions are right in proportion as they tend to promote happiness'. 'This action tends to promote happiness' contains the information that the speaker has a certain thinkable result of the action in mind – happiness-promotion – and that there might be a shortfall in the scope of the process imagined to produce that result. The rightness of the action would in this interpretation vary with two variables: the amount of happiness the complete process is thought of as producing, as well as the degree of completion of the process. So, there is a possible version of act-utilitarianism according to which X and Y in *Case One* each acts wrongly because each performs an act that tends to cause my death.

Would this be the most plausible version of act-utilitarianism? Why, someone might ask, should the fact that an act merely tends to cause an event that harms someone make the act wrong if it actually does not cause more harm? We cannot answer this question by appealing to the agent's expectations, e.g. by claiming that the agent cannot be certain that such an act will not make a difference to how harmed the victim will be. The issues discussed here all concern the question of what makes an act *objectively* wrong.

But my claim is not that act-utilitarianism necessarily must be understood in this way. Nor is it that act-utilitarianism is the correct moral view. My point is that if we want to argue that membership in a set of actions of the sort described can make an act wrong, we have to appeal to wrong-making features of the mentioned kind. Since we cannot say that the wrong-making feature is that the act harms me, or that the act is a member in a set that harms me, we have to assume that what makes the act wrong is the feature that makes it qualified for membership. What the acts have in common in these cases is that

each act tends to cause the harmful event in question. That is how they are related to the event that makes us morally concerned. In other words, to reject TSM because of this kind of thought-experiment is not to be committed to the view that mere participation can matter, but to a view about the moral relevance of tendencies to cause harm.[27]

Suppose we agree with Parfit that an act can be wrong when it is a member in the smallest set of acts such that if none of them had occurred, someone would not have been harmed. The upshot of what I have said in this section is that the mere membership in that set is not what makes it wrong. The wrong-making feature must be the one that makes the act qualify for membership. That feature is not necessarily a fact about 'the effects of the particular act', understood as the events the act marginally contributes to – the events that would not have occurred unless the act had been performed. Still, it might be a fact about the act's effects in a wider sense, such as the fact that the act initiates a causal tendency towards an event that harms someone.

First, this means that the wrong-making feature is a matter of causal relations (in a wide sense) between the act and the harmful event that unites the set. Therefore, we can admit that an act might be wrong because it belongs to the mentioned kind of set without assuming that mere participation in the set can have moral worth. Second, it shows

[27] A defence of the moral plausibility of the view I assign to those who reject TSM (owing to Parfit's thought experiments) would require a more detailed examination than I can offer here. So do several important questions about how the details should be worked out, e.g. when it comes to weighing tendencies against other sorts of effects. I believe, though, that there are moral intuitions in favour of treating tendencies to cause events that harm as morally relevant. The following case was presented to me in this journal's referee report: X kills me and Y kills me. Apart from being sufficient to kill me, X's act also saves someone else's life. Assume that what makes X's act wrong in *Case One* is that it tends to cause an event that harms me. In that case, we have to say that X's act has this wrong-making feature in the new case as well. This must then be weighed against the good of saving someone's life. The referee finds this counterintuitive and at odds with consequentialism. However, this way of reasoning seems intuitively plausible to me, as well as compatible with consequentialism. X actually does the right thing, since there is no better option available to him. But surely, even though I will die anyway, it would have been wrong of him to act in a way that tended to kill me if he could have saved your life without creating that tendency. A feature of an act is wrong-making when it makes the act wrong, other things being equal. Right acts can have wrong-making features, since other things are not always equal. For a consequentialist, the feature that makes an act wrong in every case is the property of having fewer good consequences than any other available option. On that view, X's act in the present case is right (period) – not 'partly wrong' or something like that. But this should not prevent the consequentialist from admitting that in this case, by tending to cause an event that harms, X's right act nevertheless has a wrong-making or bad feature. That his act tends to kill me *would* have made it wrong for him to act as he does if he had also been given the option of performing an act without that feature, other things being equal.

that there is another sense in which 'its effects' and 'the effects of the particular act' may come apart.

V. RANDOM SETS

From the denial of TSM, Parfit concludes that there are cases such that '[o]n any plausible theory, even if each of us harms no one, we can be acting wrongly if we together harm people'.[28] So, when I want to act rightly '[i]t is not enough to ask "Will my act harm other people?" ... I should ask, "Will my act be one of a set of acts that will *together* harm people?" '[29] I may then find that my action is wrong because it will belong to a set that has bad effects, even though the effects of my individual action are morally neutral. Among 'the countless actual cases of this kind', Parfit mentions 'pollution, congestion, depletion, inflation, unemployment, a recession, over-fishing, over-farming, soil-erosion, famine, and overpopulation'.[30]

Clear-cut and salient two-person cases of overdetermination and pre-emption like the ones discussed earlier are probably rare. The practical importance of Parfit's principles hangs on their applicability to actual cases of the sort mentioned.

Consider the collapse of the cod stock in Newfoundland some years ago. It is not unlikely that this environmental harm was overdetermined. Each person fishing for cod could then say that he did no harm, since the cod would have disappeared just as rapidly even if he had abstained. This might well be true. If we do not want to accept it as an excuse, we need some way of explaining how his act could be wrong anyway. Following Parfit, we then might think that his act was wrong because it belonged to a set of acts that together did harm. On what grounds do we place his action in the relevant set? It does not belong to that set in virtue of being performed by a fisherman, by a member of the fishermen's union, or by a member in some other formal or informal group. The relevant set of actions is random, in the sense that we do not have to assume that the agents have anything in common besides being linked to the harmful event which makes us morally concerned in the first place.

I guess that most actual cases of the kind Parfit mentions concern sets of actions that are random in this sense. Therefore, if we want to blame the particular agent because of his participation, we need a criterion of membership that does not rely on any bonds between the members in the set. Unless we have evidence for assuming that the

[28] Parfit, *Reasons and Persons*, p. 70.
[29] Ibid., p. 86.
[30] Ibid., p. 444.

particular fisherman has performed any acts that at least tended (in the sense made clear above) to contribute to the extinction, we have no reason for accusing him of having participated at all. That is an independent argument for the type of approach I have suggested above. What makes his act wrong is not just that he is a member of a certain group of agents, but the feature of his act that makes him qualified for membership. That feature is what ties him to the relevant effect.

VI. NOT ONLY SETS OF ACTIONS

Suppose I eat bad salmon. Thereafter I suffer an unrelated cerebral haemorrhage, which is serious enough to kill me exactly at the time botulin poisoning alone would have ended my life. To determine whether the cerebral haemorrhage harms me in C6's sense, we have to compare what would have happened without that haemorrhage. Since the same harmful event would have hit me anyway, my cerebral haemorrhage does not harm me, and the same way of reasoning applies to my eating salmon. So, neither botulin poisoning nor cerebral haemorrhage harms me. The comparison cannot say more than that absent botulin and haemorrhage, the harmful event would not have occurred.

Nevertheless I guess we would say that cerebral haemorrhage and botulin poisoning are both bad for me. They belong to a set of events, which are such that if none of them had occurred, I would not have been harmed. Furthermore, they belong to that set in virtue of the causal tendencies they initiate. Each event has effects such that I will die from them at that time, even if the other event does not occur. This is what makes each event in the set bad for me. In other words, there are events other than actions to which the sort of evaluative thinking discussed above can be applied.

So, TSM is not a mistake exclusive for evaluations of sets of actions performed by groups of agents. It is a mistake we can commit when attempting to evaluate the harmful or benefiting impact of any kinds of sets of events. This adds to the impression that what is at stake here is not the worth of mere participation. There is no plausible sense in which the haemorrhage and the poisoning can be said to participate in anything.

VII. IMPERCEPTIBLE EFFECTS

Are there cases in which we would say that an action belongs to a set of acts that harm people without there being *any* causal connection between the act and the harmful events in question? Parfit's 'Harmless Torturers' might appear to be an example of that kind. Consider one of

the 1,000 victims in this example.[31] He begins the day in stage 1 – mild pain. Each of the 1,000 torturers affects his pain so that in stage 1,001, he is in extreme pain. However, he would not be able to perceive the differences brought about by the individual torturers: 999 simultaneous changes would not feel less than 1,000 changes, 998 not less than 999, and so on. If they took it in turns to affect him, he would not be able to discern any stage from the immediately preceding stage. Most people seem to agree with Parfit that this confirms that an individual action can be wrong, even if the isolated effects of the individual action are imperceptible. Why is such an action wrong?

One possible explanation, favoured by Parfit in *Reasons and Persons*, is that its individual effect does harm to the victim – the torture victim's state gets imperceptibly worse for each act. In that case, the act is wrong because of its marginal contribution to physiological changes in the victim, changes that constitute harm even though they cannot be perceived by the victim. Personally, I find this explanation to be most plausible. I see no reason why we shouldn't allow changes such as 'being brought closer to the perceivable pain threshold', or 'being exposed to greater risk of feeling pain', to be labelled 'harmful' just because the victim cannot perceive them.

For the sake of argument, suppose that we nevertheless reserve 'harm' for effects that can be perceived by the person harmed. Then there is another explanation, which Parfit prefers in his 'Comments' – the act is wrong since it belongs to a set of actions which together cause harm. In that case, the example is yet another reason against committing TSM. But just as in the overdetermination cases, I think we should ask what makes each torturer's individual act a member of the set that harms the victim. To put it somewhat demagogically, is his act a member of the set that harms because it is performed by one of the torturers, or is that man a torturer because he performs one of the acts that contribute to harm?

To begin with, even if the physiological change that the act brings about really is imperceptible in the sense that the victim exposed to it cannot feel it, there are strong reasons to doubt that its contribution to perceptible harm is undetectable. Even from the victim's point of view, the overall perceptible increase in pain might be traced to the imperceptible effects of the individual acts.

Shrader-Frechette has defended 'the claim that there are no imperceptible changes in the degree of pain and no imperceptible (i.e., nonmeasurable) harms and benefits'.[32] (Note that she interprets

[31] Ibid., p. 80.

[32] Kristin Shrader-Frechette, 'Parfit and Mistakes in Moral Mathematics', *Ethics* 98 (1987), p. 58.

'perceptible' as 'measurable' here.) Gracely has suggested a probabilistic elaboration of her point.[33] In the torture case, the effect can be understood, they argue, either in terms of measurable physiological changes bringing the torture victim closer to a perceptible threshold, or in terms of measurable changes affecting the probability of feeling pain. So, although the isolated effect of each act may be imperceptible (impossible to feel by the victim), there can be detectable causal links between the act, and some harm that it causes together with other acts. Shrader-Frechette is inclined to think that 'the effects of every non-mental act are capable of being known in some way, at least on the molecular level through sophisticated instrumentation'.[34]

Again for the sake of argument, suppose that this is wrong, and that there are some cases in which it is impossible in principle to detect the causal links between an act and the harm it contributes to. This would not imply that there are no such links, or that we cannot have any good reasons for assuming that there are.

Sven-Ove Hansson argues that effects can make an act wrong even if these effects are undetectable.[35] An action has an individually detectable effect on the subject if it is possible for someone to observe how the action affects the subject, e.g. physiologically. This implies both that it is possible to observe the change, and that it is possible to link this change with the action. If it is empirically impossible to do that observation, we may still have good reasons to assume links between the action and some victim's future harm. To begin with, individually undetectable effects may be collectively detectable. Suppose, for example, that a company dumps a small amount of toxic waste in a lake. Some years later 5 per cent of the people living close to the lake develop a certain form of cancer, which occurs in 3 per cent of the average population. Biochemistry also provides us with reasons for assuming that this increase is no coincidence, but can be linked with the action. In that case, it may be impossible to say whether one specific individual is a victim of the company's action or not, although we have reasons to think that some of the cases are caused by this behaviour. Concerning an identified cancer victim, we can, perhaps, at most assume that there is a 40 per cent probability that his or her cancer has been caused by the toxic waste.

Hansson's main point is that collectively undetectable effects could also be morally relevant. This is a brief version of his argument.[36]

[33] E. J. Gracely, 'Comment on Shrader-Frechette's "Parfit and Mistakes in Moral Mathematics" ', *Ethics* 100 (1989), p. 157.
[34] Ibid.
[35] Sven Ove Hansson, 'The Moral Significance of Undetectable Effects', *Risk: Health, Safety and Environment* 101 (1999), p. 102.
[36] Ibid., p. 107.

Assume that there are biochemical theories and laboratory evidence giving us reasons to believe that the mentioned kind of toxic waste leads to a rise in the incidence of a certain form of cancer from 10.0 to 10.5 per cent. That is within the normal statistical variation of this cancer form in the average population. Then, the effects of the toxic waste would be undetectable even on the collective level. Now, assume that the company had a choice. If they had not manipulated the chemical before they dumped it, it would instead have given rise to cases of an extremely rare form of cancer. Exactly as many cases as before would have been caused, and each case would have exactly the same symptoms, treatment, and other consequences. If they had abstained from manipulation, we would have been able to detect all individual cases of cancer that had resulted from their actions. We would hardly say that the company's manipulation would make the action less wrong.

In these latter cases, harm is certainly perceptible by the victim. And even if the effect is individually and collectively undetectable, there is some plausible scientific or other theoretical evidence linking the occurrence of harm to the action. If there were no such reasons, albeit merely theoretical, to believe that the action has had something to do with the cases of cancer, we would have no reason for picking it out in a moral argument in the first place.

There is a continuum from the detectable to the barely discoverable – some theory will always be needed to allow us to draw conclusions about causal links, even on hard empirical evidence. Another point to be made is that detectability and discoverability are practical and relative notions – they have to do with the kind of evidence we could have at the present stage of science, etc. Undetectable and undiscoverable effects might be detectable and discoverable in principle.

The important thing here is merely that even if the isolated effect of an act is imperceptible, and the causal link between an act and the perceptible harm it contributes to is undetectable, we may have good reasons to assume that there is such a link. These are the sorts of reasons we might have for picking it out as a member in the set of acts that together cause harm.

Just as in the overdetermination cases, this way of reasoning could be applied to any kind of events that together cause perceptible harm, although the isolated effect of each particular event is imperceptible. TSM is in this case a mistake that arises out of unclear views about how to estimate whether an event contributes to harm. The reason that we are more interested in this kind of mistake in connection with sets of actions performed by groups of agents is merely that these are the cases where the issue of moral wrongness becomes relevant. But we can explain why it may be wrong to perform a certain action even if this particular act harms no one (because its isolated effects are

imperceptible) without assuming that mere participation matters. It may be wrong because the effects of the action may be such that it belongs to a set of events that together cause perceptible harm.

VIII. JACKSON'S DEFENCE OF TSM

Real-life assignments of rightness and wrongness are supposed to guide, prescribe, deter, or advise. Such guidance will affect people on the basis of their expectations. Jackson and others have doubted that the sort of cases Parfit discusses raise any philosophical problems at all, from the real-life subjective viewpoint. If *Case One* were to be made realistic, we would have several reasons to condemn both X and Y in terms of expected marginal contribution to harm. As Jackson points out, none of them could then be completely certain that his or her shot will not be necessary to cause my death. And if they got together before the shooting and planned the thing together, each person's contribution to this planning would constitute a marginal contribution to my death.[37]

When it comes to the case of imperceptible effects, Bart Gruzalski thinks that, from the subjective viewpoint,

> it is not clear that a case like that of the Harmless Torturers poses any problem whatsoever. If there is a foreseeable possibility that my act may make a difference to the victims' pains, and in any realistic version of such an example that seems likely, then it may be that my act would be wrong on the foreseeable account because of its negative yet expected desirability.[38]

This seems plausible. Suppose the real-life murderer Y says 'Since X also does it, my shooting will make no difference, so it cannot be wrong'. Or that a real-life torturer claims 'Since my contribution is so small, no one will feel the difference, so it cannot be wrong'. We would simply dismiss their excuses as resting upon false premisses. Y could not know that her shooting makes no difference, and the torturer cannot know that the victim will not feel his contribution. We would probably also assume that their participation in these groups in itself affected the behaviour of the other members, and this would be an indirect marginal contribution to the overall effect.

Nevertheless, as our intuitions about the fishermen who contributed to the extinction of the cod supply indicate, it would be an exaggeration to assume that there are no real-life cases in which the premisses of such excuses are true. It is quite possible that the effect (on the extinction of the cod supply) of each individual act of fishing was imperceptible. No one might have perceived the effect (in terms of extinction) of this particular act, or been able to detect its particular

[37] Jackson, 'Which Effects?', pp. 50–1.
[38] Bart Gruzalski, 'Parfit' s Impact on Utilitarianism', *Ethics* 96 (1986), p. 782.

influence on the aggregated effect of all such acts. It is also quite possible that the extinction of the cod supply was overdetermined. Therefore, the crew on each trawler may have thought that their over-fishing causes no harm, since the same harmful event will occur without their individual act. As Parfit rightly claims, over-fishing is just one out of many real-life cases in which people use the sort of excuse mentioned above. The question of whether these very common sorts of excuses can be *valid* may, after all, be of practical importance.

Apart from questioning the practical importance of TSM, Jackson argues that it is not a mistake that consequentialists should avoid. Ben Eggleston defends Parfit against Jackson's arguments.[39] Although I agree with Eggleston that Parfit's rejection of TSM withstands Jackson's arguments, my reasons are somewhat different.

Jackson's first objection draws upon an alleged similarity between *Case One*, *Case Two* and *Case Three*. Jackson notes that in *Case Three*, Parfit claims that Y acts rightly, 'because Y benefits someone else but does not harm me'.[40]

> *Case Three*. As before, X tricks me into drinking poison of a kind that causes a painful death within a few minutes. Y knows that he can save *your* life if he acts in a way whose inevitable side-effect is my immediate and painless death. Because Y also knows that I am about to die painfully, Y acts in this way.[41]

Now Jackson argues that there can be no relevant difference between *Case Two* and *Case Three* when it comes to the wrongness of Y's act. This is because he thinks that Y benefits someone in both cases – me (because I am spared some pain) in *Case Two* and someone else in *Case Three*. The question of who gets the benefit cannot be relevant, at least not from a consequentialist point of view. But then we should also conclude that in *Case One*, Y's action is neither wrong nor right, since she harms me no more than in *Case Two*, according to Jackson.

However, this argument stems from misreading Parfit. To begin with, Parfit assumes that although Y's act in *Case Two* 'is in one way slightly worse for me, since it shortens my life with a few minutes', this is 'outweighed by the fact that Y saves me from a painful death'.[42] As the case is described, I am therefore neither benefited nor harmed by the effects of Y's specific act. It is on a par with X's act when it comes to harming me. So, one relevant difference between the cases is that no one is benefited in *Case Two*, while someone is in *Case Three*.

[39] Eggleston, 'Should Consequentialists Make Parfit's Second Mistake?'.
[40] Jackson, 'Which Effects?', p. 47.
[41] Parfit, *Reasons and Persons*, p. 71.
[42] Ibid.

As Eggleston points out, Jackson also ignores another relevant difference. In *Case Three*, Y's act is dependent upon X's, while it is independent in *Case Two*. We know that in both cases, had Y not fired, I would have died anyway, and in that case, I would not have died unless X had poisoned me. It is only in *Case Three*, however, that I would not have died unless X had poisoned me. In that situation, Y kills me because I am about to die anyway. In this case X's specific act, but not Y's, harms me in C6's sense, since I would not have died unless X had poisoned me. (A third feature that might be considered relevant is that Parfit describes my death as a side-effect of Y's act in *Case Three*. But my guess is that one cannot avoid appealing to intuitions about *subjective* wrongness in order to make that assumption relevant.) In all three cases each act suffices to harm me, but though this is the crucial feature of the first two case descriptions, the third case contains other important facts about the effects of the acts – Y's act causes a benefit, and X is causally responsible for Y's act of causing my death. These are the sort of facts that affect Parfit's delimitation of the set of wrong actions.

Jackson's second objection is that those who hold that X and Y both act wrongly in *Case One* introduce agent-relative values, and that 'runs counter to the whole thrust of consequentialist thinking'.[43] If we hold that X ought not to shoot, we assign relevance to the mere difference in who brings my death about, because if 'X shoots, he joins in with Y in killing me; if X refrains from shooting, it is all done by Y', and that is the only relevant difference. One way of dealing with this objection, according to Eggleston, is to assume that although X does not contribute to the agent-neutral disvalue of there being more killings in the world, he contributes to some other agent-neutral disvalue, such as there being more violence in the world. But that solution seems incompatible with Parfit's point, and it could hardly ground the conclusion that TSM is a mistake. That there is more violence in the world is a direct effect of X's specific act.

Another objection of Eggleston's is that Jackson equates two distinct senses of 'relevance' in his argument. 'One is *relevance to the question of which outcome is better*. We may call this *outcome-relevance*. Another is *relevance to the moral assessment of agents' conduct*. We may call this *conduct-relevance*'.[44] Eggleston's point is that while consequentialists should regard the question of who does the killing as outcome-irrelevant, they need not regard it as conduct-irrelevant. He uses one of Jackson's examples, in which I can choose between killing one of your parents and letting you kill both. In that case, my conduct

[43] Jackson, 'Which Effects?', p. 47.
[44] Eggleston, 'Should Consequentialists Make Parfit's Second Mistake?', p. 8.

is better from a consequentialist point of view if I pull the trigger. Hence, consequentialists can admit that who does the killing is conduct-relevant and think that this matters for the wrongness of X's conduct in *Case One* as well.

I do not find that inference convincing. It is correct that from a straightforward consequentialist point of view, my conduct is better if I shoot one of your parents than if I let you shoot both. However, it is not better in virtue of the fact that I am the one pulling the trigger, but because my pulling the trigger has fewer bad consequences in agent-neutral terms (than the only other alternative action left open to me in the thought experiment). Jackson's point is not affected – consequentialists should assess the worth of conduct in terms of its contribution to agent-neutral value. And in *Case One*, X's joining in makes no difference in those terms.

Eggleston also claims that in this second objection, Jackson simply assigns to consequentialists a view of the relation between the good and the right that is question-begging. Why shouldn't consequentialists be 'free to say, as Parfit does, that an act is wrong if it is one of a *set* of acts that together have bad consequences'?[45] I am inclined to agree. If there are cases in which a contribution to agent-neutral value is an effect of a set of acts, and in which it is impossible to trace the elements in this overall effect to the individual acts, why should consequentialists have to refrain from condemning the individual's participation in this set? However, it seems unlikely that a consequentialist should find an individual act wrong *solely* in virtue of the agent's participation in that group of agents. The consequentialist has no reason to worry about X's shooting me unless he sees some kind of connection between this shooting and the promotion of agent-neutral values. As before, my suggestion is that the consequentialist could condemn each act in *Case One* because it is a sufficient cause of my death. This is a fact about the consequences of each act, about their tendency to cause harm. These features are what make each action a member of the set of acts that together harm.

The third and final objection Jackson presents focuses the difference between genuine overdetermination and pre-emption. Now Jackson assumes that Parfit neither could nor would claim that Y acts wrongly in a case where his killing me is pre-empted by X's killing me. In such a case, 'Y would not even be a partial cause of something bad that happens to me. But this conclusion means that Parfit must hold that the fine detail of what happens inside me is absolutely crucial in a way which is hard to believe'. And he goes on to say that deontologists might

[45] Ibid., p. 7.

assume that such details matter, but that it is hard to accommodate them within a consequentialist framework.[46]

It is somewhat surprising that Jackson thinks that this is what Parfit would say. Take *Case Two*, in which Parfit says that although X does not kill me (because his killing me is pre-empted by Y's killing me a few minutes before I would have died from X's action), X acts wrongly. Clearly, Parfit assigns no importance to the difference between pre-emption and simultaneous overdetermination. In this argument, Jackson shifts between two views of how the harmfulness of an action should be estimated. If the harmfulness of an act is equal to the harmfulness of the events it causes, X's act will not be harmful in *Case Two*. That way of estimating harmfulness is not, though, what Parfit has in mind. According to the comparative way of estimating harmfulness stated in Parfit's principle C6, there is no crucial difference between pre-emption and simultaneous overdetermination. In both *Case One* and *Case Two*, it is true of each action that had it not been performed, the same amount of harm would have occurred anyway, so neither action harms me. It is also true of both cases that this amount of harm would not have occurred if none of the acts had been performed, and this is the sort of fact that makes both acts members of the set that harms. So, it seems to me that Parfit's principle produces exactly the result Jackson would like to see. The fine detail about which bullet strikes my heart first is irrelevant.

IX. MERE PARTICIPATION AND THE SHARE-OF-THE-TOTAL VIEW

Even if we suppose with Eggleston that to deny TSM is to accept that mere participation can matter morally, we can deny TSM without being forced to accept 'the Share-of-the-Total View', because the latter claims more than that mere participation can matter. The Share-of-the-Total View, as I understand it, says that the worth of producing a certain harm/benefit should be distributed equally among the individual contributors to it (if there are no significant differences between them with regard to the production of the benefit).[47] A consequence of the Share-of-the-Total View is that there can be cases in which mere participation in a group can render a member blame or credit for an effect, even though his participation makes no difference to the production of this effect.

However, 'Mere Participation Can Matter' does not imply the Share-of-the-Total-View. Wrongness/rightness owing to participation does not

[46] Jackson, 'Which Effects?', p. 49.
[47] Parfit, *Reasons and Persons*, p. 68.

have to be equal for all members, and it does not exclude that worth owing to degree of marginal contribution also adds to the act's total worth. So, the tension between Parfit's denying the Share-of-the-Total-View and his rebuttal of TSM is not genuine incoherence, even if avoidance of TSM should imply that mere participation can matter. There could be other arguments against the Share-of-the-Total View than that it assigns worth to mere participation.

Eggleston argues, though, that Parfit's indictment of the mistake of believing the Share-of-the-Total View implies that mere participation cannot matter. He thinks that the only element in the Share-of-the-Total View that can make it a mistake is the belief that mere participation can matter.[48] To illustrate this point, Eggleston refers to Parfit's 'First Rescue Mission' example, in which I choose between joining three people in saving a hundred miners, and saving ten other people single-handedly. If I do not join the rescue mission, a fifth person will join the other three, and these four will save the hundred miners.[49] 'My claim . . . is that this is precisely *why* Parfit regards the Share-of-the-Total View as a mistake: because it says I have a moral reason to join the others – in other words, because it claims that my participation matters'.[50]

Eggleston says initially that what can be wrong with the Share-of-the-Total-View is the belief that in such a case, my joining the others has 'a moral worth comparable to that of an act of actually saving twenty-five lives'. He adds, however, that '[i]n short, it leads me to think that I have a moral reason to join the others'.[51] And it is the latter shorter explication that grounds his conclusion. Something substantial is missing in this alleged circumlocution, though, because I may consistently believe that mere participation matters, and therefore that there is some moral worth in joining the rescue party, without believing that the moral worth of joining equals the worth of saving twenty-five lives. Nor will 'mere participation matters' force me to think that I have a stronger reason to join than to save ten lives single-handedly. Again, Parfit's rejection of the Share-of-the-Total View does not imply that mere participation cannot matter.

X. CONCLUSION

Parfit says that even if an act harms no one, this act may be wrong because it is one of a set of acts that together harm other people.

[48] Eggleston, 'Does Participation Matter?', p. 98.
[49] Parfit, *Reasons and Persons*, p. 67–8.
[50] Eggleston, 'Does Participation Matter?', p. 97.
[51] Ibid., p. 98.

According to a common reading, this implies directly that mere participation can matter. However, if we focus on what makes an act a member of that kind of set, we see that there is another possibility. An act may be wrong when it belongs to such a set because it belongs to the set in virtue of having certain effects. Although those effects may not marginally contribute to harm, we may nevertheless think that they make the act wrong. The set of acts that together harm other people in the overdetermination case is such that if none of the acts had been performed, these other people would not have been harmed. What makes a certain act a member of that set is the effects of the act. Each act is a member in virtue of the causal tendency it initiates towards the event that has attracted our attention. In the case of imperceptible effects, what the acts have in common is that we have some evidence for assuming that they contribute causally to perceptible harm.

I do not claim that we have to reject TSM or embrace C7. What I claim is that we can do so without assigning worth to mere participation. The question of whether mere participation can matter must be settled on other grounds than those Parfit appeals to in his rejection of TSM.[52]

bjorn.petersson@fil.lu.se

[52] Wlodek Rabinowicz gave me many useful comments on an earlier version of this article. I am also indebted to the Practical Philosophy Seminar in Lund, as well as the referee. This article was prepared as part of a research project for the Swedish Science Council.