

Defence Against Parfit's Torturers

Jeff McMahan

C15.S1

1. Torturers, Harmful and Harmless

C15.P1 One of Derek Parfit's concerns in the section of *Reasons and Persons* on 'mistakes in moral mathematics' is whether there can be imperceptible harms and benefits.¹ My concern here is quite different: it is with what it can be permissible to do in defence against people who will otherwise inflict a perceptible, though quite small, harm on each of a large number of victims. Although our topics are quite different, there are nevertheless two ways in which Parfit's discussion provides the basis for mine. First, slightly modified versions of his core examples are ideal for illustrating the issues I will discuss. Second, each of the rival explanations he distinguishes of why it can be wrong to cause imperceptible increases in a person's pain might also partly explain why it might be permissible to engage in harmful defensive action against a person who threatens to cause a barely perceptible harm to each of a large number of victims. The summary of Parfit's discussion in this first section will therefore serve to introduce my own discussion.

C15.P2 Parfit invites his readers to consider a group of people I will call

C15.P3 *The Traditional Torturers*

Each of a thousand people has a button that, if pressed, will cause an innocent person to experience a thousand distinct but simultaneous electric shocks for eight hours. Each shock on its own would be imperceptible but a thousand such shocks together constitute torture. Each of these thousand Traditional Torturers presses his button, thereby causing a thousand victims each to suffer agony for eight hours.²

I have been presenting and gradually refining the material in this essay in lectures given in various places since 2013. I have been benefited from comments from audiences at the University of Leeds, the Rutgers-Camden School of Law, the University of Manchester, the University of Stockholm, the University of Oslo, Boston College, and the University of Southern California. I am also greatly indebted for written comments on earlier drafts to Ben Bronner, Derek Parfit, Victor Tadros, and Patrick Tomlin, and for illuminating discussion to Fiona Clarke, Johann Frick, Helen Frowe, and, especially, Tim Campbell, Shelly Kagan, Frances Kamm, and Larry Temkin.

¹ Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1987 repr.), ch. 3, sects. 28 and 29.

² Parfit, *Reasons and Persons*, 80. My statements of Parfit's examples are modified in several trivial ways.

Jeff McMahan, *Defence Against Parfit's Torturers* In *Principles and Persons: The Legacy of Derek Parfit*. Edited by: Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan, Oxford University Press (2021).

© Oxford University Press. DOI: 10.1093/oso/9780192893994.003.0016

C15.P4 It is obvious that what each Traditional Torturer does is wrong. And the obvious explanation is that, although each shock on its own is imperceptible, ‘the *total effect* of what each torturer does’ is to inflict great suffering on a single victim.³

C15.P5 Parfit next asks us to consider

C15.P6 *The Harmless Torturers*

Each of a thousand people has a button that, if pressed, will administer an imperceptible electric shock to each of a thousand innocent people for eight hours. Each Harmless Torturer’s thousand victims are the same individuals as every other Harmless Torturer’s victims. (We can suppose, indeed, that the thousand button pressers and the thousand recipients of shocks in this case are the same people as those in the corresponding roles in the case of the Traditional Torturers.) All the Harmless Torturers press their buttons at the same time. The result is that the thousand victims each suffer agony for eight hours.

C15.P7 There are, I will assume, no morally relevant differences among the Harmless Torturers. All are equally culpable, none would be more seriously harmed by being killed than any other, and so on.

C15.P8 Parfit claims, correctly in my view, that each Harmless Torturer is ‘acting just as wrongly as’ each Traditional Torturer. Yet none of the Harmless Torturers on his own causes a perceptible effect on any one of the victims. Parfit suggests two explanations of how what the Harmless Torturers do is nevertheless wrong. The first presupposes that one’s pain or suffering can be made imperceptibly worse and that one can be harmed by an imperceptible increase in one’s pain or suffering. In that case, Parfit says, ‘since each torturer adds to the suffering of a thousand victims, each torturer imposes a *great total sum* of suffering.’⁴ This explanation is similar to that which applies to the action of the Traditional Torturers. The difference is that the total sum of suffering inflicted by each Traditional Torturer is concentrated in one period in the experience of one victim, whereas the total inflicted by each Harmless Torturer is dispersed equally among a thousand victims.

C15.P9 Parfit concedes, however, that many people believe that any increase or worsening of pain or suffering must be perceptible. If this belief is correct, no Harmless Torturer worsens the condition of, or harms, anyone. Thus, Parfit writes, ‘if we cannot appeal to the effects of what each torturer does, we must appeal to what the torturers together do. Even if none of them causes any pain, they together impose great suffering on a thousand victims.’⁵ This is the second explanation of why what the Harmless Torturers do is wrong.

³ Parfit, *Reasons and Persons*, 80; italics added.

⁴ Parfit, *Reasons and Persons*, 80; italics added.

⁵ Parfit, *Reasons and Persons*, 80.

C15.P10 Parfit next presents a third example:

C15.P11 *The Single Torturer*

One person has a button that, if pressed, will administer an imperceptible shock to each of a thousand innocent people for eight hours. This person knows that each of these thousand people is about to experience eight hours of pain from natural causes that will be equivalent in intensity to the aggregate pain from 999 of the individually imperceptible shocks. If the Single Torturer presses his button, the result will thus be the same as it would be if he were adding his shocks to those of 999 other Harmless Torturers—that is, all thousand victims will experience the same pain that the victims suffer in the case of the Harmless Torturers. The Single Torturer presses his button.⁶

C15.P12 In the first printing of *Reasons and Persons*, Parfit indicates that he believes that the Single Torturer acts wrongly and that the first explanation of why the Harmless Torturers act wrongly provides a plausible explanation of why the Single Torturer acts wrongly as well. He acts wrongly because each of the thousand imperceptible increases in pain he causes harms, or worsens the condition of, the victim of that increase. He is, therefore, 'imposing... a great total sum of suffering' on his thousand victims.

C15.P13 In this same first printing, however, Parfit concedes that many people believe that the Single Torturer does not act wrongly. These people, he writes,

C15.P14 believe that it cannot be wrong for someone to affect others in a certain way, if this person knows both (1) that these effects will be imperceptible, and (2) that they will not be part of a set of effects that, together, are perceptible. Since this belief is widely held, and not implausible, it is better not to appeal to the effects of what each torturer does. Even if we believe that there can be imperceptible harms and benefits, as I do, it is better to appeal to what groups together do. This appeal is less controversial.⁷

C15.P15 In short, if the fact that the Single Torturer's act causes a great total sum of suffering is insufficient to make his act wrong, the fact that each Harmless Torturer's act causes a great total sum of suffering cannot be the explanation of why his act is wrong. The relevant difference between the acts of the Harmless Torturers and

⁶ Parfit's example is in Parfit, *Reasons and Persons*, 81. My statement of the example differs from his in two ways. First, the thousand victims in his example have already been suffering pain for some unspecified period when the Single Torturer presses his button. Second, the suffering they are already experiencing is equivalent to the sum of only five hundred of the individually imperceptible pains. Their suffering is thus only half as bad as that of the victims of the Harmless Torturers.

⁷ Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984), 82.

the act of the Single Torturer is that the effects of each Harmless Torturer's act are 'part of a set of effects that, together, are perceptible', whereas that is not true of the effects of the Single Torturer's act. What Parfit must mean here by 'effects' is 'effects that are the products of *agency*'. The imperceptible effects caused by a single Harmless Torturer are wrong because they are additions to other effects of the same sort caused by other Harmless Torturers with whom he is acting together, as a group, or with whom he is collaborating. Although the imperceptible effects caused by the Single Torturer are additions to other effects of the same sort, these other effects are not caused by people with whom the Single Torturer is collaborating. Understood in this way, Parfit's second explanation of why what the Harmless Torturers do is wrong applies to their acts but not to the act of the Single Torturer. Thus, if we believe that the Single Torturer does not act wrongly, we can accept the second explanation of why the Harmless Torturers act wrongly, even though we cannot accept the first.

C15.P16 In a later reprinting of *Reasons and Persons*, however, Parfit slightly modified his view. There he writes that

C15.P17 I believe that the Single Torturer *is* acting wrongly. How can it make a moral difference whether he produces bad effects jointly with other agents, or with Nature? I therefore prefer, in both cases, to appeal to the effects of single acts. Some people disagree. Even if we believe that there can be imperceptible harms and benefits, it may thus be better to appeal to what groups together do. This is less controversial.⁸

C15.P18 In this passage, Parfit expresses scepticism about the view that adding bad effects to those produced by others with whom one is collaborating is wrong in a way that adding equivalent bad effects to equally bad effects produced by natural causes is not. But if there really is no moral difference between these two ways of adding to bad effects from other sources, then both the first and second of Parfit's explanations of the wrongness of each Harmless Torturer's act also imply that the Single Torturer's act is wrong as well, and presumably equally wrong. But Parfit nevertheless suggests that, because an explanation that appeals to the distinct wrongness of collaborative harming implies that the acts of the Harmless Torturers are wrong without implying that the act of the Single Torturer is wrong, it is perhaps better, because less controversial, to adopt that explanation rather than the one that appeals to the effects of single acts.

C15.P19 I think, however, that Parfit is right to doubt that there is a distinctive wrongness to collaborative harming that is absent from acts that add or contribute to harms from sources other than the acts of collaborators. I will try to reinforce his scepticism by appealing to some further examples in Section 3. If I am right, we

⁸ *Reasons and Persons* (1987 repr.), 82.

should reject rather than embrace his second explanation of the wrongness of the acts of the Harmless Torturers. This conclusion would not have unduly troubled Parfit, for, as he indicates, his only reason for endorsing this explanation is that it is likely to be acceptable to more people than his first explanation, which is the explanation that he favours on its merits. I think, however, that the first explanation is also inadequate, for reasons I will present in Section 2. The best explanation is a variant of Parfit's second explanation, purged of the collectivist component of which Parfit was rightly sceptical.

C15.S2 2. Why, and to What Extent, a Harmless Torturer's Act Is Wrong

C15.P20 My concern in this essay is, as I noted, not with whether there can be imperceptible harms and benefits or whether it can be wrong to cause effects that are imperceptible. In the remainder of this essay, therefore, I will assume that the individual shocks inflicted by the Traditional Torturers, the Harmless Torturers, and the Single Torturer all cause perceptible, though only barely perceptible, pains. Even though it lasts for eight hours, each of these pains is only the mildest, almost negligible, annoyance. But, as with the imperceptible effects in Parfit's original examples, a thousand simultaneous annoyances amount to agony. (On the assumption that the pains they cause are perceptible, the Harmless Torturers are not, strictly speaking, harmless. But I will continue to use this label nonetheless.)

C15.P21 When Parfit's examples are understood in this way, his first suggested explanation of why what the Harmless Torturers do is wrong clearly applies. Because each inflicts a thousand perceptible pains, each inflicts a great total sum of pain, or suffering. The question remains, however, whether this best explains why what they do is wrong, or whether it explains why what they do is as seriously wrong as it is—that is, as wrong as what the Traditional Torturers do.

C15.P22 We can, I believe, see that this explanation does not account for how seriously wrong what the Harmless Torturers do is by considering a further group of people who also inflict tiny harms on a large number of people.

C15.P23 *The Inflictors*

Each of a thousand people has a button that, if pressed, will administer a barely perceptible electric shock to each of a thousand innocent people for eight hours. There is, however, no overlap among each of these people's victims. All of these Inflictors of tiny harms press their buttons at the same time. The result is that each of a million innocent people experiences a barely perceptible pain for eight hours.⁹

⁹ This example first appears in my review of Helen Frowe's book, *Defensive Killing*, in *Ethics* 126 (2016): 825–31, at 826. The discussion there anticipates some of what I say in this section.

C15.P24 What the thousand Inflictors do is wrong. But it is not as seriously wrong as what the Traditional Torturers do, and thus not as seriously wrong as what the Harmless Torturers do. Yet in certain respects what each Inflictor does is the same as what each Harmless Torturer does: each inflicts a barely perceptible pain for eight hours on each of a thousand innocent people. And each Inflictor causes the same total sum of suffering that each Harmless Torturer causes. The Inflictors might, moreover, act in collaboration, as a group, so that the pains caused by each are ‘part of a set of effects’—a million tiny pains—that they together produce. And each Inflictor might be just as culpable as each Harmless Torturer (each might, for example, mistakenly believe that she *is* a Harmless Torturer).¹⁰ But it would still not be true that each Inflictor’s act is as wrong as that of a Harmless Torturer. Indeed, even if we were to stipulate that each Inflictor has five thousand victims, and so causes a total sum of suffering that is five times as great as that caused by each Harmless Torturer, it would still be difficult to believe that her act is as wrong as that of a Harmless Torturer. So Parfit’s first explanation that appeals to the total sum of harm that an act causes cannot account for the intuitive difference between what the Inflictors do and what the Harmless Torturers do. What, then, is the correct explanation?

C15.P25 The degree to which what these different people do is wrong is, I believe, a function not only of the magnitude of the total harm they each cause, and of the magnitude of the harms they each cause to particular individuals, but also of the magnitude of the harms to particular individuals to which they each *contribute*. Thus, whereas each Inflictor does not contribute to a greater harm than that which she alone inflicts, each Harmless Torturer contributes to *torture*—or, more precisely, to the experience of eight hours of agony by a thousand individuals. It is, I think, the fact that each Harmless Torturer’s act makes a contribution to very *great harms to individuals*, and to *very many* individuals, that makes the acts of the Harmless Torturers more seriously wrong than those of the Inflictors. And this same fact makes the Single Torturer’s act more seriously wrong as well (on the assumption, one will recall, that the bad effects caused by the Inflictors and the Single Torturer are *perceptible* pains).

C15.P26 There can be groups of people who inflict many tiny harms that are intermediate between the Harmless Torturers and the Inflictors. One example is

C15.P27 *The Contributors*

Each of a thousand people has a button that, if pressed, will deliver two simultaneous, barely perceptible shocks to each of five hundred innocent

¹⁰ Parfit, writing in 1984, referred to all his torturers using male-gendered pronouns. For consistency, I have done the same. Because it is important in this essay to distinguish clearly between what each individual of a certain type does and what all the individuals of a type together do, I do not use plural pronouns to refer to individuals. For variety, I use female-gendered pronouns to refer to Inflictors, to those I will call ‘Contributors’, and to third-party defenders.

people for eight hours. There are two thousand potential victims. All the Contributors press their buttons. Two hundred and fifty of them inflict shocks on five hundred of the victims, another two hundred and fifty inflict shocks on a different five hundred, and so on. The result is that each of the two thousand victims suffers five hundred simultaneous tiny pains for eight hours.

C15.P28 Each Contributor inflicts the same total sum of suffering that the Single Torturer and each Traditional Torturer, Harmless Torturer, and Inflictor causes. But the pain that each Contributor inflicts on each of her five hundred victims is twice as intense, and therefore presumably twice as bad, as that which any of the others inflicts on any individual victim. This makes what each Contributor does in one important respect more seriously wrong. And each Contributor's act is in another respect more seriously wrong than that of each Inflictor: for each Contributor adds her two simultaneous pains to 498 equivalent pains that her victims suffer at the same time from other sources, whereas the lesser pains caused by each Inflictor are not contributions to greater pain. Each Contributor's act is, however, less seriously wrong, at least in this one respect, than the act of a Harmless Torturer; for the pains each Contributor inflicts are contributions to suffering by individuals that is only half as intense as that to which each Harmless Torturer contributes.

C15.P29 Given these various similarities and differences between the effects of the Contributors' acts and the effects of the acts of the other inflictors of tiny harms, it is perhaps difficult to determine whether what each Contributor does is as seriously wrong as what each Harmless Torturer does. There may be reasonable disagreement about whether one of the considerations just noted is more important than another, and thus reasonable disagreement about whether the act of a Contributor is more or less seriously wrong than the act of a Harmless Torturer. In my view, however, it is not implausible to suppose that what each Contributor does is less seriously wrong, at least slightly. Although each Contributor contributes more to the suffering of each of her victims, it may be more important to the determination of how seriously wrong her act is that it contributes to the suffering of victims whose suffering is only half as bad as that of the victims of the Harmless Torturers.

C15.S3

3. The Irrelevance of Collaboration

C15.P30 My claim, then, is that it is a crucial part of the explanation of why what each Harmless Torturer does is wrong, and seriously wrong, that his act contributes to, or increases, individual suffering (or, more generally, harm) that is *extremely bad*. It is also important, of course, that it does this to a very large number of victims.

The great suffering to which each Harmless Torturer contributes is, of course, the product of the acts of the other Harmless Torturers. Each one's act is wrong, in other words, because it is a contribution to what they together do, which is to cause great suffering to a large number of victims. It may therefore seem that my claim is no different from Parfit's second explanation. But, as I indicated earlier, I share Parfit's scepticism, stated in the later reprinting of *Reasons and Persons*, about the moral significance of collaboration. What matters, in my view, is that each Harmless Torturer is knowingly and intentionally adding another tiny pain to the very great suffering that each victim experiences from other sources. With Parfit, I believe that it does not matter to the wrongness of the Harmless Torturer's act whether those sources are agents with whom he is collaborating or something else. I believe, for example, that there is no moral difference between what a Harmless Torturer does and what the Single Torturer does. If the Single Torturer knows that each of a thousand people is about to experience 999 simultaneous insect bites, each of which will cause pain of the same intensity and duration as the pain from a single Harmless Torturer's shock, and if the Single Torturer chooses then to press his button, what he does is just as wrong, and for the same reasons, as what a Harmless Torturer does.

C15.P31 Yet I know of other moral philosophers who believe that it is essential to the evaluation of each Harmless Torturer, and to understanding what may permissibly be done to him (for example, in defence, as punishment, or to compel him to compensate his victims), that he is collaborating with others in the infliction of wrongful harms. Some may believe this because they believe that collaboration can affect the degree of a wrongdoer's culpability. Others may believe it because they believe that, when one collaborates with others, one bears some responsibility for what those others do.¹¹ There are obvious ways in which this latter belief can be true. If, for example, each Harmless Torturer is motivated to press his button in part by pledges from the others that they will also press theirs at the same time, or if his conscience is silenced by the thought that a great many others will also do exactly what he is tempted to do, then each Harmless Torturer does indeed bear some responsibility for the acts of all the others.

C15.P32 Yet we can assume, or stipulate, that nothing of this sort occurs in the case of the Harmless Torturers. Let us assume, in the remainder of this essay, that there is an independent reason why each Harmless Torturer wants the thousand victims to experience pain at a specific time. Perhaps each torturer hates the victims because they together constitute a group to which the torturer is averse. And suppose that there is some precise time at which the victims plan to celebrate an event that is of significance to them as members of the group. It is *then* that each Harmless Torturer wants them to suffer. So each Harmless Torturer is strongly motivated entirely independently of the others to inflict as much pain on the

¹¹ This is Frances Kamm's view, expressed to me in discussion.

victims as he can at that exact time. Of course, each can inflict only a tiny pain on each victim. So each is pleased to discover that the others share his antipathy and plan to do exactly what he plans to do. Each is delighted to learn that, because of the others, he will not just be causing a tiny pain to each victim but will be contributing to each victim's *agony*. Perhaps they all rejoice in this together and in this way are collaborators. But each would have pressed his button in the absence of the collaboration and each *will* press his button at the relevant time whatever the others might do. Given these assumptions, it does not seem that any one Harmless Torturer bears responsibility for what any of the others does.

C15.P33 Another problem for the view that collaboration in the infliction of wrongful harm is a distinct wrong that makes what each Harmless Torturer does more seriously wrong than what the Single Torturer does is that it may be unclear what exactly collaboration is, or what it involves. An examination of some examples in the grey area supports the view that collaboration is morally irrelevant, apart from any bearing it might have on an agent's culpability, or any influence it might have on what people actually do. We might, for example, reflect on these variations on the case of the Harmless Torturers.

C15.P34 *The Aspiring Collaborator*

A man knows that 999 people will, at a precise time, collaborate in causing agony for eight hours to a thousand innocent victims by each inflicting a tiny pain on each victim. This man is not known to any of the collaborators and cannot communicate with them but he is able to inflict an equivalent tiny pain on each of the thousand victims at the same time that the 999 inflict theirs, and does so.

The Hopeful Torturers

Each of a thousand people has a button that, if pressed, will administer a barely perceptible electric shock to each of a thousand innocent people for eight hours. Each presses his button at the same time in the hope, based on reports he has heard, that many others will also inflict shocks on these same victims at that same time. None of these people can communicate with any of the others; indeed, none knows the identity of any of the others or even whether anyone other than himself has the ability to inflict shocks on the same thousand victims. Each one nevertheless makes a tiny contribution to the suffering of agony by a thousand victims.

C15.P35 It seems to me that, although the Aspiring Collaborator would like to collaborate with the others, he does not do so. Yet it seems clear that what he does is no less wrong than what each Harmless Torturer does, and that he is no less culpable than any of those who do collaborate. One could, of course, contend that there is a sense in which he does act together with the others even though they do not know it, or a sense in which he is a member of the group that together inflicts the

great suffering even though the others do not know he is. But one must then ask: what difference does this make, morally? What matters is not the proper description of his relation to the others but that he deliberately makes a contribution to the suffering of agony by a thousand innocent people.

C15.P36 Similar remarks apply to the Hopeful Torturers. It seems to me that, although they would like to, they do not collaborate or act together. They are not a group and certainly do not constitute a collective agent. But again this seems not to matter. Whether they satisfy some condition for acting together seems irrelevant. What matters is that each one successfully tries to make a contribution to the suffering of agony by a thousand innocent victims. Hence each one's act is as seriously wrong as that of an individual Harmless Torturer. This is compatible with the possibility that each individual Hopeful Torturer is less culpable than each Harmless Torturer. In the absence of collaboration, each Hopeful Torturer cannot act with the assurance that he will contribute to the infliction of agony on the thousand victims. For all he knows, his act may be like that of an Inflictor rather than that of a Harmless Torturer. If this might diminish his culpability, it illustrates the way in which collaboration may be relevant to culpability.

C15.P37 If it is correct that collaboration is irrelevant apart from its effects on the motivations and acts of those who collaborate, this supports the claim that there is no morally significant difference between any one Harmless Torturer and the Single Torturer. Each deliberately adds his tiny pains to the agonizing pains that he knows his victims will suffer simultaneously from causes that are independent of his action. My claim that the wrongness of what each Harmless Torturer does is partly explained by the magnitude of the individual suffering to which he contributes can thus be restated in prioritarian terms. According to prioritarianism, any increase or decrease in an individual's well-being matters more the worse off that individual is at the time. (There is also a version of prioritarianism that is concerned with lifelong well-being but it is not relevant here.) Although the pain that any Harmless Torturer inflicts on any victim is tiny, each victim is, at the time of the individual Harmless Torturer's action, extremely badly off in absolute terms because of the action of the other 999. This is the core element of the explanation of why each Harmless Torturer's act is so much more seriously wrong than each Inflictor's act. For each of the victims of any of the Inflictors is, we are implicitly assuming, reasonably well off, in the way most people normally are, at the time the Inflictor inflicts his tiny pains. The prioritarian view is also the core part of the explanation of why the act of the Single Torturer is as seriously wrong as that of any Harmless Torturer.¹²

¹² I should be clear that I am not here embracing any particular version of prioritarianism. I am simply appealing to a consideration that is an element of different versions of the larger view—namely, that the worse off an individual is independently of one's action, the more seriously objectionable it is to inflict on that individual a fixed amount of pain, if other things are equal.

C15.S4

4. Liability to Defensive Harming

C15.P38 The considerations that explain why an act of harming is wrong may also help to explain why the person who will otherwise do the act may be morally liable to be harmed in defence of the potential victim or victims of the act. And these same considerations also contribute to determining *how much* defensive harm the person may be liable to suffer. The degree of harm to which he is liable is a matter of what is known as *narrow proportionality*. Narrow proportionality is a constraint on a *liability justification* for harming a person. If the degree of harm inflicted on a person as a means of preventing the harm he will otherwise inflict on others exceeds the amount that can be justified on grounds of moral liability, that harm is disproportionate in the *narrow sense* and, according to many writers on defensive harming, wrongs the potential wrongdoer. Narrow proportionality contrasts with *wide proportionality*, which is a constraint on a *lesser-evil justification* for the infliction of harm to which the victim is not liable.¹³

C15.P39 I have thus far discussed four types of agent: Traditional Torturers, Harmless Torturers (including the Single Torturer), Inflictors, and Contributors. I will refer to agents of all four types as 'harmers'. I have also noted four factors that I believe contribute to determining the degree to which what these various harmers do is wrong. These are

- C15.P40 (1) the magnitude of the harm that each inflicts on individual victims,
- C15.P41 (2) the total sum of harm that each inflicts (which is a function of the magnitude of the harm to individual victims and the number of those victims),
- C15.P42 (3) the magnitude of the total harm to each victim to which the harmer contributes, and
- C15.P43 (4) the degree of the harmer's responsibility or culpability.

C15.P44 One might question whether the degree of a harmer's responsibility or culpability can affect the degree to which that harmer's action is wrong. The wrongness of an act and the culpability of the agent are in general distinct; hence, an act can be seriously wrong and yet the one who does it may not be culpable at all. This is what is meant when it is said that a person is *excused* for wrongful action. Yet I believe it is not unreasonable to suppose that the wrongness of an act—even the objective or fact-relative wrongness of that act—is exacerbated if the agent is highly culpable rather than merely responsible or altogether excused in the doing of the act. I will, in any event, assume that this is so, though it is not essential to the points I will make.

¹³ For elucidation and further discussion, see Jeff McMahan, 'Proportionate Defence' (revised and expanded version), in Jens Ohlin, Larry May, and Claire Finkelstein (eds.), *Weighing Lives in War* (Oxford: Oxford University Press, 2017): 131–54.

- C15.P45 The same four factors all seem relevant as well to the degree of harm to which the different types of harmer may be liable in defence of their victims. In one of the cases I have discussed—the Traditional Torturers—the amount of harm that each harmer inflicts on his individual victim is very great: agony for eight hours. This, combined with the assumption that each Traditional Torturer is culpable, makes each one liable to a defensive harm significantly greater than that which he would otherwise inflict. Let us assume that each is morally liable to be killed to prevent him from torturing his single victim. This, I believe, is the common-sense view and is not implausible.
- C15.P46 By contrast with a Traditional Torturer, each harmer of the other three types inflicts no more than a very slight harm on any individual victim. Yet we may assume that each of these harmers is culpable to some degree. Each, we may suppose, acts maliciously without justification or excuse. If these were the only relevant factors, these harmers might be liable to a defensive harm somewhat, though not substantially, greater than the slight harm they would inflict on their victims. But of course all three of these types of harmer have a great many individual victims. Because of this, the total sum of harm that each inflicts is, as Parfit notes, as great as that which a Traditional Torturer inflicts. This seems to increase the amount of harm to which each is liable as a matter of defence. Yet the amount of harm to which an individual Inflictor is liable to prevent her from pressing her button seems intuitively not to be very great. I believe, for example, that it would be disproportionate to cause her to suffer a *hundred* simultaneous pains, each of the same intensity and duration as of one of the *thousand* pains she would otherwise inflict. It seems, therefore, that the total sum of harm that an act inflicts is not a source of liability to substantial defensive harm, provided the act neither causes nor contributes to a significant harm to any individual.
- C15.P47 Although no Harmless Torturer causes more than a trivial harm to any one individual victim, each does *contribute*, albeit in a comparatively trivial way, to the infliction of great harm on each of his victims. Because of this, the total sum of harm caused by each Harmless Torturer has greater significance than that caused by each Inflictor, though less significance than that caused by a Traditional Torturer. But, again, the total sum of harm that a person causes seems to have little significance except insofar as it is the product of significant harms to many individuals or of contributions to significant harms to many individuals. This seems true of the significance that the total sum of harm has both in determining the degree to which an act of harming is wrong and in determining the amount of harm to which the harmer is liable to prevent him from harming others.
- C15.P48 Assuming, at least for the sake of argument, that each Traditional Torturer is liable to be killed if that is necessary to prevent him from pressing his button, and assuming that each Inflictor is liable only to a comparatively small harm to prevent her from pressing her button, we can now ask how much harm a Harmless Torturer is liable to as a means of preventing him from pressing his button. This, I will

argue, depends on and varies with certain wholly *circumstantial conditions of liability*. The circumstantial conditions of liability are conditions over which a potentially liable person has no control. They contrast with the *agential conditions of liability*, such as what the person intends and why she intends to do it, over which she does have some control. In the case of the Harmless Torturers, one relevant circumstantial condition is how many other Harmless Torturers will simultaneously be prevented from pressing their buttons, either by the same defending agent or by others. If the only possibility is preventing one Harmless Torturer from pressing his button, this Harmless Torturer may be liable to only a relatively small defensive harm—one greater than that to which an Inflictor is liable but substantially smaller than that to which a Traditional Torturer is liable. It might, for example, be proportionate to break the Harmless Torturer's fingers, or his wrists, as a means of preventing him from pressing his button; though because the prevention of his contribution alone would make very little difference to the magnitude of any individual's suffering, even that might be disproportionate in the narrow sense. Certainly, at least in my view, it would be disproportionate to kill him.

C15.P49 Suppose next that five hundred of the Harmless Torturers have already been defensively incapacitated. The only possibility now is to prevent one more from pressing his button. The amount of harm to which this Harmless Torturer is liable is, I think, less than that to which he would be liable if the five hundred others had not been incapacitated. This is because, while the magnitude of the harm he would inflict on each of the thousand victims is the same in each case, the magnitude of the harm to each individual *to which he would contribute* when only 499 others are contributing is less than half the magnitude of that to which he would contribute if all 999 others were contributing as well. If 999 of the Harmless Torturers had already been incapacitated, the remaining one would be liable to the same amount of harm to which an individual Inflictor is liable. And the amount of harm to which an Inflictor is liable does not depend on how many others would also be prevented from pressing their buttons. The one remaining Harmless Torturer would have effectively become a mere Inflictor rather than a contributor to torture.

C15.P50 There is another way in which the amount of defensive harm to which a Harmless Torturer is liable might vary. Suppose that in conditions in which it is impossible to prevent 999 of the Harmless Torturers from pressing their buttons, the maximum amount of harm that it would be proportionate to inflict on the remaining Harmless Torturer to prevent him from pressing his button is n . Next suppose that it becomes possible to prevent two hundred of the thousand Harmless Torturers from pressing their buttons, though only by inflicting on each of them a harm several times greater than n ($n \times 3$). It seems plausible to suppose that each of the two hundred Harmless Torturers could be liable to that amount of harm—that is, that it would not be disproportionate to inflict a harm of $n \times 3$

on each. This is because the elimination of the contribution of any one Harmless Torturer matters more when it is part of a substantial reduction of great suffering than when it occurs alone or is part of only a relatively insignificant reduction of great suffering.

- C15.P51 Suppose next that 750 of the Harmless Torturers have already been incapacitated. The only further defensive option is to prevent two hundred more from pressing their buttons. This can be done only by inflicting a harm of $n \times 3$ on each of them. We are assuming that $n \times 3$ is the maximum harm to which each of two hundred Harmless Torturers could be liable as a means of reducing the victims' suffering from agony (a thousand simultaneous small pains) to great but not agonizing suffering (eight hundred simultaneous small pains). Given this assumption, I think it would be disproportionate to inflict $n \times 3$ on each of two hundred Harmless Torturers when that would reduce their victims' suffering by the same amount, though from moderate suffering (250 simultaneous small pains) to very mild suffering (fifty simultaneous small pains). This is because the elimination of the contribution of any one Harmless Torturer matters more when it is a part of a substantial reduction of very great suffering than when it is part of an equally substantial reduction of suffering that is significantly less bad.

C15.S5

5. Defensive Killing

- C15.P52 I have suggested that the prevention of one Harmless Torturer's contribution to the suffering of the thousand victims matters more when it would be part of a substantial reduction in the suffering of many victims, each of whom would otherwise experience very great suffering. The prevention of one Harmless Torturer's contribution therefore matters most when it would be accompanied by the prevention of the contributions of all the other 999. In this case, the harm to which a Harmless Torturer would be liable should be greater than that to which he would be liable if the prevention of his contribution were part of a less substantial reduction of suffering that was already less bad.
- C15.P53 Suppose that a third party has only two options: she can either do nothing, thereby allowing all thousand Harmless Torturers to press their buttons, or she can simultaneously painlessly kill all of them, thereby preventing them from pressing their buttons. Given the assumption that it would be permissible for a third party to kill all thousand Traditional Torturers, I think we should accept that it would be permissible, in these conditions, for a third party to kill all thousand Harmless Torturers. Indeed, if we accept that it would be obligatory for a third party to kill all thousand Traditional Torturers, provided that this could be done without excessive personal cost, then I think we should accept, with the same proviso, that it would also be obligatory for a third party to kill all the Harmless Torturers rather than not kill any. For the morally relevant effects would

be the same in both cases: the same thousand culpable people would be killed to prevent them from doing what would make them causally and morally responsible for the agonizing suffering of the same thousand innocent people.

C15.P54 The justification for the killing of each Traditional Torturer is a liability justification. Each is morally liable to be killed because killing him is a necessary and proportionate means of preventing him from torturing an innocent person. Similarly, in a choice between killing all the Harmless Torturers and allowing them to press their buttons, the justification for killing them is also, I believe, a liability justification. Assuming that it is in fact permissible to kill them, one might argue that the justification must be a liability justification by default, for there does not seem to be any other form of justification that applies: the Harmless Torturers do not deserve to die, killing them is not the lesser evil (as eight hours of agony is not worse than death, when death involves the loss of many years of good life), the third party (we may assume) is not specially related to any of the victims, and so on.

C15.P55 Some philosophers with whom I have discussed the case of the Harmless Torturers say, however, that they think that no Harmless Torturer is liable to be killed, irrespective of how many it is possible for a third party to kill. This is because each Harmless Torturer, like each Inflictor, causes no more than a trivial pain to any individual. Although each Harmless Torturer, again like each Inflictor, causes a great many such pains, these pains do not aggregate to become morally significant. To kill a Harmless Torturer to prevent him from inflicting these trivial pains would therefore, these philosophers claim, be disproportionate.

C15.P56 This understanding of the grounds of liability raises a problem to which Philip Pettit has called attention in another context. Pettit observes that

C15.P57 the failure to impose a regime of corporate responsibility can expose individuals to a perverse incentive. Let human beings operate outside such a regime, and they will be able to incorporate, so as to achieve a certain bad and self-serving effect, while arranging things so that none of them can be held fully responsible for what is done.¹⁴

C15.P58 Applied to Parfit's cases, the concern is that the claim that no Harmless Torturer is liable to be killed has a highly implausible implication—namely, that the Traditional Torturers, each of whom is probably liable to be killed and certainly liable to a defensive harm significantly greater than eight hours of agony, could achieve exactly the same bad effects while altogether escaping liability to serious defensive harm simply by reprogramming their buttons to operate in a way that would make them Harmless Torturers rather than Traditional Torturers.

¹⁴ Philip Pettit, 'Responsibility Incorporated', *Ethics* 117 (2007): 171–201, at 196.

- C15.P59 This is not just a concern about the practical consequences of the acceptance of a view that would exempt the Harmless Torturers from liability to serious harm. It is an objection to the truth of the view itself. But is the implication actually unacceptable? It is clearly possible for a person to arrange his causal environment in such a way as to enable him to engage in serious wrongdoing while ensuring that others have no proportionate means, and indeed no permissible means, of preventing his action—for example, by surrounding himself with a large number of ‘innocent shields’, thereby making it impossible for others to prevent him from committing a murder without killing all of his innocent captives as a side effect.
- C15.P60 There is, however, an important difference between the example of the murderer with many innocent shields and the case of the Harmless Torturers. Defensive action that would kill the murderer’s innocent shields would be disproportionate in the wide sense. The murderer is himself liable to be killed but the liability justification for killing him is overridden by the rights of others who would be killed if one were to act on that justification. But the claim of those who deny that any Harmless Torturer is liable to be killed is that killing any of them would be disproportionate in the narrow sense. Because harm that is disproportionate in the narrow sense is, by definition, harm in excess of that to which a person is liable, no one can be liable to harm that is disproportionate in that sense. The question, therefore, is simply whether a Harmless Torturer can be liable to be killed in the circumstances.
- C15.P61 It is important, in considering this question, not to confuse the issue of liability with that of desert. It is obvious that, even if each Harmless Torturer deserves some degree of harm, the harm of death would be vastly disproportionate as a matter of desert. But the harm to which a person can be liable is sensitive to morally arbitrary circumstances (again, the ‘circumstantial conditions of liability’) in ways that the harm a person deserves is not. In particular, the harm to which a threatening person can be liable as a matter of defence is highly sensitive to who else will unavoidably be harmed, and by how much, if the threatening person is not harmed. The harm to which a person is liable is thus determined by comparisons among possible courses of action, whereas the harm a person deserves depends only on what that person has done or is doing. Because of this, the harm to which a person can be liable may be much greater or much less than the harm, if any, that he deserves.
- C15.P62 Because of this comparative dimension to liability, there is good reason to accept that, if the choice is between killing all and killing none, each of the Harmless Torturers is morally liable to be killed. In the circumstances, it is unavoidable that either a thousand potential victims will each suffer a grave harm or a thousand different people—the Harmless Torturers—will each suffer a significantly greater harm. But none of the potential victims bears any responsibility for this situation. Each of them is entirely morally innocent. Each Harmless Torturer, by contrast, is fully responsible for a small part of the grave harm that

each of the potential victims may suffer. Because each Harmless Torturer would be an intentional contributor to these grave harms, each is culpable and bears a share of the responsibility for the fact that a third party must choose between inflicting a greater harm on each of them and allowing them together to inflict grave harms on each of the potential victims. Even though death is a significantly greater harm than eight hours of agony, it is not a disproportionate harm in these circumstances, in which each Harmless Torturer could have chosen, without personal cost, not to press his button, thereby avoiding becoming culpably responsible for a situation in which it is unavoidable that each of a thousand people will suffer a very great harm. It is thus a matter of justice that each of the innocent people should be defended from being caused to suffer a very great harm even if each of the culpably responsible people must be caused to suffer an even greater harm as a means of defence.

C15.P63 It is essential to each Harmless Torturer's liability to be killed in certain conditions of choice not only that the instances of individual suffering to which he intentionally contributes are extremely severe but also that he contributes to the suffering of many victims. We can imagine a case in which this latter feature is absent.

C15.P64 *The Harmless Torturers with a Single Victim*

Each of a thousand people has a button that, if pressed, will administer a tiny electric shock to a single innocent person for eight hours. Each knows that 999 other people will press a button at the same time that will have the same effect on the same victim. The result of their combined acts will be that a single victim will experience agonizing suffering for eight hours.

C15.P65 Suppose that the only way that the potential victim or a third party can prevent any of these Harmless Torturers from pressing their buttons is to kill them.

C15.P66 Although some moral philosophers with whom I have discussed this example believe that it would be permissible for the potential victim or a third party to kill any or all of these Harmless Torturers, I believe that killing any number of them would be disproportionate and therefore that none of them is morally liable to be killed.¹⁵ Killing only one of them would make only a trivial difference to the suf-

¹⁵ For a defence of a view that implies that, if a Traditional Torturer is liable to be killed, any or all of the Harmless Torturers with a single victim are liable to be killed, see Helen Frowe, *Defensive Killing* (Oxford: Oxford University Press, 2014), 78 and 175. For objections to this view, see my review of the book, cited in n. 9. Frowe's view combines the idea that collaboration is itself morally significant with the idea that the defensive harm to which a threatener is liable is affected by the magnitude of the harm to individuals to which he *contributes*. According to her view, the defensive harm to which a threatener is liable is, when he acts alone, a function of the magnitude of the harm he would himself cause, whereas when he acts in collaboration with others, the defensive harm to which he is liable is instead a function of the magnitude of the harm that he and his collaborators would together cause, to which he would only contribute. This view implies that all the Harmless Torturers, and even all the

fering of the victim but would, we may suppose, deprive the Harmless Torturer of as many as thirty to fifty years of good life. Killing a second one as well would also make only a trivial difference and the suffering that would be reduced by the second killing would be slightly less bad than that which would be reduced by killing the first. Killing all of them would, of course, prevent the innocent victim from suffering eight hours of agony, but the cost of that would be that a thousand people, each of whom would otherwise have inflicted only a tiny pain on one person, would together be deprived of a total of *30,000–50,000 years* of good life. I find it impossible to believe that this could be proportionate.

C15.P67 If I am right about this, this example is an instance in which culpable people can evade liability to defensive harm by acting together in the way identified by Pettit. Suppose that each of these thousand Harmless Torturers wants the innocent victim to suffer agony for eight hours. Each would act on his own as a Traditional Torturer to make this happen, were that necessary. But each understands that if he were to attempt to inflict the eight hours of agony by himself, he would become liable to be killed, as killing him would be a necessary and proportionate means of defending the victim. Given that each knows this, and that each knows that there are 999 others who would also be willing to be a Traditional Torturer to make the victim suffer, these thousand people decide to achieve their shared aim by becoming Harmless Torturers, thereby, I believe, avoiding becoming liable to be killed.

C15.P68 I have told this story in a way that makes these Harmless Torturers collaborators. And the form of their collaboration has clear moral significance, in that it is intended by each to enable him and the others to achieve an aim they share that is seriously wrong, and to do so without exposing themselves to liability to the only form of defensive action that is possible in the circumstances. Because of this, each is arguably more culpable, and in principle liable to greater harm, than each would be if they were all only Hopeful Torturers with a single victim. Still, because each of these Harmless Torturers with a single victim is willing to be a Traditional Torturer, they are all highly culpable even in the absence of the collaboration. But given that all that each intends to do is to inflict a single barely perceptible pain for eight hours on a single innocent person (rather than single-handedly inflict eight hours of extreme agony on that person), it remains impossible for me to believe that it could be proportionate to kill any of them in defence of the victim.

C15.P69 It is worth mentioning, if only in passing, that Pettit's implied suggestion that the problem he cites can be avoided by appealing to a doctrine of collective responsibility is of no help in understanding what harm these Harmless Torturers with a single victim might be liable to suffer as a means of defending their victim.

Harmless Torturers with a single victim, are liable to be killed, but the Single Torturer (who, like the Harmless Torturers, has a thousand victims) is not.

It may well be that, because of their planning and coordination, they together constitute a group agent. And there is certainly a sense in which, were they to act in coordination, they would be collectively responsible for the agony their victim would suffer. But we cannot, I think, infer anything about their individual liability from these claims that is not already implied by the fact that each acts in collaboration with the others to achieve an immoral aim that they share in a way that enables each to avoid becoming morally liable to be killed. Recognizing that they constitute a group agent that is responsible for torturing an innocent person would not make it any less implausible to suppose that any of them are liable to be killed.

C15.P70 Similarly, claims about group agency and collective responsibility are unlikely to help in understanding how much harm the Harmless Torturers with a thousand victims might be liable to suffer in defence of those victims. As I hope the subsequent discussion will show, each Harmless Torturer's liability can vary with the options available to the defender, and there seems to be scope for reasonable disagreement about individual liability in many of the possible cases. And because I have stipulated that, although there is a sense in which the Harmless Torturers collaborate, what each does is causally independent of what any other does, claims about group agency and collective responsibility are substantially less plausible in this case. In these conditions, I suggested, each Harmless Torturer is relevantly like the Single Torturer, who simply adds his pains to many others with causes (such as insect bites) that are entirely beyond his control. There can be no collective responsibility in such a case.

C15.S6

6. Liability in Varying Conditions of Choice

C15.P71 In Section 5 I sought to explain why, when a third-party defender can only kill all the Harmless Torturers or else kill none, every Harmless Torturer is liable to be killed. Perhaps the most salient circumstantial condition of their liability is that *either all* the innocent victims who bear no responsibility for the unavoidability of great harm *must* suffer grave harms *or all* the Harmless Torturers who are culpably responsible for the unavoidability of great harm *must* suffer even graver harms. In these conditions, proportionality is not assessed by reference to the magnitude of the individual harms that each Harmless Torturer threatens to inflict. The assessment of proportionality instead requires weighing the difference in moral responsibility between the Harmless Torturers and the innocent victims against the difference in the magnitudes of the individual harms that the members of each group might suffer, given that the members of one group must suffer those harms. On the assumption that it is proportionate to kill a Traditional Torturer rather than allow him to inflict eight hours of agony on an innocent victim, it seems that it is also proportionate to kill the Harmless Torturers rather than allow them to inflict eight hours of agony on their innocent victims.

- C15.P72 The assessment of liability is quite different, however, in conditions in which harm is unavoidable but can be divided or shared between those who are morally responsible for this and those who are not. Suppose, for example, that a third party could either kill a Traditional Torturer, thereby preventing him from causing any pain to his victim, or cause him to be permanently quadriplegic, thereby preventing him from inflicting nine hundred of the thousand tiny pains he would otherwise inflict on his victim. In the latter case, the victim would have to suffer a hundred tiny pains—relatively mild but still significant pain—for eight hours. Assuming that continued life in a quadriplegic condition is very substantially better than death, it might be that in these conditions of choice the Traditional Torturer would not be liable to be killed but only to being made quadriplegic.
- C15.P73 This is not, however, a matter of proportionality but is instead a matter of a different constraint on a liability justification—namely, necessity. Although it is generally assumed that the necessity constraint simply requires that a defender cause the least harm necessary to achieve a particular defensive aim, I believe that this is an oversimplification. Suppose, for example, that a third party has two options for defending an innocent victim from a Traditional Torturer. She can prevent the Traditional Torturer from inflicting all thousand tiny pains, but only by killing him; or she can prevent him from inflicting 999 of the tiny pains by giving him a sharp pinch. If she had only the first option, the killing would be both necessary and proportionate. But given that she has the second option, killing the Traditional Torturer is unnecessary. Killing him is of course *physically* necessary for the full defence of the victim, but it is *morally* unnecessary because necessity is sensitive to trade-offs between the defensive harm inflicted and the wrongful harm to be prevented. The same is true, though less obviously so, when a third party can choose between killing a Traditional Torturer to spare the victim all thousand tiny pains and causing the Traditional Torturer to become quadriplegic to spare the victim nine hundred of the tiny pains. The first of these options is wrong because it is morally unnecessary.
- C15.P74 I believe, moreover, that a person cannot be liable to defensive harm that is morally unnecessary—or, as this claim is sometimes expressed, that necessity is ‘internal’ to liability.¹⁶ If this is true, and if killing the Traditional Torturer is unnecessary when he could be caused to be quadriplegic instead, it follows that in these conditions he is not morally liable to be killed. He is liable only to be made quadriplegic and his innocent victim is morally required to endure a hundred tiny pains for eight hours (though not as a matter of *liability*). This may seem unfair to the victim but in fact the requirements of necessity and proportionality often require innocent victims to endure harms from culpable aggressors (for

¹⁶ For a defence of this claim, see Jeff McMahan, ‘The Limits of Self-Defence’, in Christian Coons and Michael Weber (eds.), *The Ethics of Self-Defence* (New York: Oxford University Press, 2016): 185–210, at 195–7.

example, when the only way to prevent a malicious person from giving an innocent person a vicious pinch is to kill him).

C15.P75 Harms can be divided or shared by culpable threateners and innocent victims either individually or collectively. Suppose that a third party could (1) kill every Traditional Torturer, thereby preventing each from inflicting any pain on his intended victim, (2) cause each Traditional Torturer to become quadriplegic, thereby preventing him from inflicting nine hundred of the thousand tiny pains he would otherwise inflict, or (3) allow all the Traditional Torturers to inflict agony on their victims. This is a case in which individual sharing of harms is possible. Either every victim but no Traditional Torturers will be harmed, or each torturer but no victims will be harmed, or each torturer and each victim will suffer a lesser harm than he or she would suffer in one of the other possible outcomes. In these conditions, in which *some* harm is unavoidable, the third option offers a way for each victim and each torturer to suffer a lesser harm than either would suffer in one of the other options, thereby sharing the unavoidable harm, though of course they do not literally divide the *same* harm between them.

C15.P76 Suppose next that a third party could (1) kill all of the Traditional Torturers, (2) kill five hundred of them, or (3) kill none of them, thus allowing each to inflict agony on his victim. Suppose the third party chooses, impermissibly in my view, to kill only five hundred of the Traditional Torturers. No individual torturer is forced to share a harm with his victim. Rather, the five hundred who are not killed suffer no harm while their victims each suffer all the harm that, in conditions in which individual sharing was possible, each might have shared with her torturer. But there is a collective sharing of harms. Half the members of the group of victims are spared any harm while each of the Traditional Torturers suffers the full harm that, in other conditions, he might have shared with his victim.

C15.P77 In the remainder of this section, I will discuss only cases in which collective sharing of harms is possible. In particular, I will assume that the only means of preventing any Harmless Torturer from pressing his button is to kill him, but that a third-party defender has, in various possible cases, options other than simply killing all or killing none. In these conditions, the liabilities of the Harmless Torturers can be quite complicated. This is an important difference between the Harmless Torturers and the Traditional Torturers (in the original example, in which the only option is to kill them)—a difference that does not emerge in Parfit's discussion, which is limited to understanding whether, and if so why and to what extent, what the Harmless Torturers do is wrong. Any Traditional Torturer is liable to be killed no matter how many others might be prevented from pressing their buttons. If a third party has a choice between killing one Traditional Torturer and killing none, it is permissible, and arguably obligatory, for her to kill the one. And, given the choice, it is always permissible, and perhaps obligatory, for a third party to kill more Traditional Torturers rather than fewer. But if a third party has a choice between killing one Harmless Torturer and killing

none, it would be impermissible for her to kill the one. Although the one Harmless Torturer would be liable to a certain amount of harm to prevent him from pressing his button, it would be disproportionate to kill him; for killing him would, in these circumstances, make almost no difference to the suffering of any victim. And because one cannot be morally liable to harm that is disproportionate in the narrow sense, no one Harmless Torturer can be liable to be killed on his own. It is, moreover, not always permissible for a third party, if given the choice, to kill more Harmless Torturers rather than fewer. Indeed, as we will see, the truth seems to be the other way around.

C15.P78 If these claims about the liability of individual Harmless Torturers in certain conditions of choice are true, it seems that at least *some* of the Traditional Torturers *could* achieve their aims, or at least have an equal probability of achieving the aims, while evading liability to be killed. They could do this by reprogramming their buttons, thereby becoming Harmless Torturers, when a defender is able to prevent them from pressing their buttons only by killing them, but can kill varying numbers of them between, and including, all thousand and none. I will give examples of this in the remainder of this section. These will be examples involving only Harmless Torturers, but one can understand them as Harmless Torturers who were formerly Traditional Torturers but have reprogrammed their buttons. In these examples, all the Harmless Torturers are *conditionally* liable to be killed. They would, for example, all be liable to be killed *if* the defender's options were restricted to killing all and killing none. And even in cases in which the defender has many options for killing varying numbers, each is initially conditionally liable to be killed because, if it is permissible for the defender to kill only some number fewer than all, it makes no difference which particular Harmless Torturers she kills. But once she kills that number, the remaining Harmless Torturers cease to be even conditionally liable to be killed.

C15.P79 Suppose next that the only options were killing five hundred Harmless Torturers and killing none. My intuition is that five hundred of them might well be liable to be killed in these circumstances—that is, that it would be proportionate to kill them. (More precisely, all thousand are conditionally liable to be killed; it would not matter, if other things are equal, which five hundred one killed.) I am, however, less confident about this judgement than I am about the judgement that all thousand are liable to be killed if the choice is between killing all and killing none. This is because, as I suggested earlier, the prevention of any one Harmless Torturer's contribution matters more—so that he is liable to greater harm—when it is part of a larger reduction in suffering that is itself greater or more intense.

C15.P80 One could continue to explore a great many other possibilities, including choices between killing some and killing none, sequential choices, and 'paired' choices, such as a choice between killing some and killing none when some others will be prevented from pressing their buttons by other causes. But there is a more

difficult question. Suppose that the only way to prevent any Harmless Torturer from pressing his button is to kill him. And suppose that a third party could, at no personal cost, kill any number of them. How many would it be permissible for her to kill? It would certainly be impermissible for her to kill all thousand rather than 999; for if 999 were killed, the remaining one would not, as I noted, be a Harmless Torturer but an Inflictor, a minor nuisance to a thousand people. Killing him in addition to killing the other 999 would be disproportionate. And the same seems to apply to killing 999 rather than 998. If 998 were killed, both the remaining ones, even acting together, would be nuisances rather than contributors to torture.

C15.P81 Surely, however, there is some number (though not a number that can be identified with precision) whose killing would make a sufficient difference to the suffering of each victim that killing them would be proportionate. Killing a hundred Harmless Torturers, for example, would substantially reduce the suffering of all thousand victims. It would still leave them in terrible pain, but the reduction would be far from trivial.

C15.P82 The problem, however, is that we are not comparing killing a hundred only with killing all thousand, or only with killing none. We are assuming that it is possible to kill any number, from zero to a thousand. We know that to kill only one is impermissible, whether the alternative is to kill none or to kill some greater number. And we know that killing all thousand is impermissible when killing somewhat fewer is an option. Is there an optimal number between zero and a thousand that it would be permissible to kill? If there were, it might be impermissible to kill more than that, or fewer.

C15.P83 There is a general problem here that the work of certain philosophers, above all that of Larry Temkin, has made familiar.¹⁷ It would, as we have seen, be wrong to kill only one Harmless Torturer. This is because the reduction in suffering it would bring to any one victim is so slight that, even though there are a thousand victims, it would be disproportionate to kill a person, even a culpable person, to bring about those barely perceptible reductions. But killing two of the thousand Harmless Torturers would seem to be even more seriously wrong than killing one. The killing of the second one would, like the killing of the first, achieve only a trivial reduction in the suffering of any victim; but it would also be wrong for an additional reason—namely, that the trivial reduction in each victim's suffering that would be achieved by killing the second Harmless Torturer would be a reduction in suffering that would be less bad, albeit only slightly, than the suffering that would be reduced by the killing of only one. That is, while the suffering reduced by killing only one Harmless Torturer would measure 1,000 in intensity, that which would be reduced by killing a second would be 999 in intensity. (This is

¹⁷ Larry Temkin, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning* (New York: Oxford University Press, 2012).

true not just when the reductions would be temporally sequential but also, as I am imagining, when they would be simultaneous. The suffering that each victim would endure if one Harmless Torturer were killed, and would thus be reduced if a second were killed as well, is 999.) So the all-things-considered reason not to kill a second Harmless Torturer is marginally stronger than the all-things-considered reason not to kill the first; thus, while killing one would be wrong, killing two would be slightly more seriously wrong. It might seem, indeed, that the extent to which killing two would be more wrong than killing one would be greater than the extent to which killing one would be more wrong than killing none.

C15.P84 There is, however, another relevant consideration. I have thus far identified two reasons why killing $x + 1$ Harmless Torturers (x being some number between 0 and 999) is more seriously wrong than killing x . These are (1) that killing the additional Harmless Torturer would make only a trivial difference to the suffering of any victim, so that there would be one more death that would produce only a trivial good effect, and (2) the suffering of each victim that would be trivially lessened by killing $x + 1$ would be slightly less bad than that which would be lessened by killing only x . A further relevant consideration is (3) that the killing of the additional Harmless Torturer in the killing of $x + 1$ would be *part of a greater reduction* in the suffering of each victim than would be achieved by killing only x .

C15.P85 It is difficult to assess the relative importance of factors 2 and 3. To test our intuitions, we might ask whether reducing the suffering that a person must endure from 600 to 590 matters more than, less than, or imprecisely just as much as reducing the suffering that a person must endure from 300 to 280. My admittedly weak intuition in this comparison is that it matters more to reduce the greater suffering by less than to reduce the lesser suffering by more. This comparison, which echoes the earlier brief discussion of the Contributors, may suggest that the prioritarian consideration—the magnitude of the individual suffering to which an individual Harmless Torturer would contribute, and thus the magnitude of the suffering that would be reduced by killing him—matters at least slightly more. Yet there may well be no general truth about the relative importance of these two factors. Suppose one could reduce the suffering that a person must endure from 1,000 to 999 or reduce that which another person must endure from 999 to 997. In this case my weak intuition is that one ought to produce the slightly greater reduction of the slightly less intense suffering.

C15.P86 The issues here are subtle. If a third party kills fewer Harmless Torturers, the intensity of the suffering that would be reduced by the killing of each one would be greater than the intensity of the suffering that would be reduced by killing more in addition. But by killing the additional ones as well, the third party would produce a greater reduction in the suffering of each victim. Whereas the first consideration favours killing fewer, the second favours killing more. I will simply assume that, in a comparison between killing a greater number and killing a

lesser number, in conditions in which one could kill any number, these opposing considerations simply cancel each other out.

C15.P87 This still leaves the fact that, in killing $x + 1$ rather than only x , one would be causing one more person to die as a means of achieving no more than a trivial reduction in any victim's suffering. In this respect, killing one more is always worse than killing one fewer. If this consideration is not outweighed by some countervailing consideration, and killing only one Harmless Torturer would be wrong (for exactly this reason), then killing two would be more wrong than killing one, killing three more wrong than killing two, ... killing 999 more seriously wrong than killing 998, and killing all thousand more seriously wrong than killing 999.

C15.P88 But we have now reached the conclusion that killing all thousand would be wrong, and not just wrong when the only alternative is killing 999. Assuming that 'more seriously wrong than' is a transitive relation, the foregoing abbreviated chain of reasoning entails that killing all thousand is wrong even when the alternative is killing none. Indeed, it seems to entail that, when it is possible to kill any number, killing *any* number is wrong; for it is always better to kill one fewer, all the way down to none. But these conclusions are, I think, clearly mistaken. If it is permissible to kill all the Traditional Torturers rather than killing none, it should also be permissible to kill all the Harmless Torturers when the alternative is killing none.

C15.P89 This is an example of what Temkin calls a 'spectrum argument'. We begin with an act that kills one person and judge that it would be wrong. We then consider a sequence of acts (a notional rather than temporal sequence) that are like the first except that each succeeding act kills one person more than is killed by the preceding act. We judge that each act that kills one more is more seriously wrong than the act that kills one fewer. An act that kills substantially more people than another should therefore be substantially more seriously wrong. But this does not seem to be true. Killing five hundred does not seem to be substantially more wrong than killing only one. Indeed, killing five hundred seems permissible—unless there is some other number that would be optimal so that one ought to choose that number rather than five hundred.

C15.P90 It is, I believe, morally unacceptable to conclude that, although killing all thousand Traditional Torturers is permissible, and better than killing any fewer than that, killing all thousand Harmless Torturers is impermissible when one could kill fewer than that, and indeed more seriously wrong than killing any lesser number. While it is true that killing all thousand would be worse than killing 999, and is impermissible when killing 999 is an option, it is not true that killing all thousand would be worse than killing none, or one, or two, or fifty. Killing all thousand would be permissible if these were the only other options.

C15.P91 It may be, as Temkin argues, that these arguments compel us to accept that various relations that we have assumed must be transitive are not in fact

transitive. This may be true, for example, of ‘more seriously wrong than’. But I will not speculate about general responses to the problems posed by spectrum arguments. I have nothing to add to Temkin’s masterly discussions of these arguments.¹⁸ I will instead return to the question I posed earlier—namely, whether, if it were possible to kill any number of the Harmless Torturers, there is some number that it would be optimal to kill. One reason to think there is not is that, as we saw, it seems that for any number one might kill, it would always be better to kill one fewer. But we must somehow resist the regress argument to the conclusion that it is not permissible to kill any. What follows is one suggestion for how we might think about the difficult problem of defence against the Harmless Torturers.

C15.P92 We should, I think, rule out the permissibility of killing only a small number of Harmless Torturers: one, or two, or ten, or probably even twenty. This would, I will say, be ‘too few’. Even though the magnitude of the individual suffering that would be reduced by killing too few would be very great, the amount by which it would be reduced would be too insignificant to justify killing people, even people who are culpable to whatever extent the Harmless Torturers are culpable. Killing too few would, moreover, be impermissible irrespective of how many other Harmless Torturers it would be possible to kill instead. It would, for example, be impermissible to kill too few rather than kill none and it would be impermissible to kill too few rather than any number greater than too few. (I do not claim to know where the boundary is between too few and enough. It is, in any case, likely to be imprecise. It might be, for example, that any number up to n is too few, and any number over $n + 10$ is enough for killing that number to be potentially permissible. But it may be indeterminate whether killing some number between n and $n + 10$ would be too few or enough.)

C15.P93 *When it would be possible to kill any number*, we should also rule out the permissibility of killing too many. In these conditions, it would be impermissible to kill all thousand, or 990, or even 950. For, as I noted earlier, killing a thousand rather than 999 would be to kill someone who, in the circumstances, would be only a trivial nuisance to a thousand people. To kill 980 rather than 950 would be to kill thirty people, each of whom would make only a tiny contribution to relatively mild pain that would be suffered by a thousand people. This too would be disproportionate. (Because we are assuming that the defensive killings would be simultaneous rather than temporally sequential, these thirty could be any of the thousand. Their identities do not matter. If one were to kill 980, at least thirty of those killed would not have been liable to be killed, as one cannot be liable to disproportionate harm. At least thirty would therefore have been wronged, but it is arbitrary which thirty these are. If by contrast, the killings were sequential,

¹⁸ Theron Pummer, however, has much to add. See his ‘Sorites on What Matters’, in Jeff McMahan et al. (eds.), *Ethics and Existence: The Legacy of Derek Parfit* (Oxford: Oxford University Press, forthcoming).

those wronged might be the thirty killed after 950 had already been killed. Those thirty would not have been liable to be killed, though only because of a morally arbitrary circumstantial condition—namely, their position in the queue.)

C15.P94 Killing too many—that is, killing some number in the upper range (which also has an imprecise or indeterminate lower boundary)—differs in at least four important ways from killing too few. First, killing too few is always wrong whatever the relevant alternative might be. But killing some number in the upper range may be permissible if only certain other options are possible. For example, whereas killing 980 is impermissible if it is possible instead to kill 960, or perhaps 940, it is permissible to kill 980 if the only alternative is to kill none, or to kill some number that would be too few (or even, perhaps, some number not too far above the upper boundary of the 'too few' range).

C15.P95 The second difference is that the numbers in the 'too many' range seem more extensive than those in the 'too few' range. This is because the magnitude of the suffering that would be reduced by killing some number in the 'too few' range is very great, so that reducing it by some fixed amount matters much more than a reduction by the same amount of suffering that is much less severe, such as the suffering that would be reduced by some of the killings if one were to kill some number of Harmless Torturers in the 'too many' range. Suppose a third party can kill either eighty of the Harmless Torturers or kill none. To kill eighty, she would have to make a significant personal sacrifice. Suppose further that she would be acting alone, so that her action would reduce the suffering of the victims from 1,000 to 920. Let us assume that she would be morally required to make the sacrifice because this would be a substantial reduction of extremely great suffering. But now suppose that she could act in conjunction with another third party who will kill eight hundred Harmless Torturers whatever she does. Even if it would be permissible for her to kill an additional eighty Harmless Torturers in these circumstances, she might not be required to make the same sacrifice to bring about a reduction of eighty when it would be a reduction of only comparatively mild suffering.

C15.P96 The third difference is that, when it is possible to kill any number of Harmless Torturers, killing some number in the 'too many' range rather than some number just outside that range is more seriously wrong than killing some number in the 'too few' range rather than some number just outside that range. This is because the suffering that would be reduced by killing some number within the 'too many' range is relatively mild, whereas that which would be reduced by killing some number within the 'too few' range is severe. Yet—and this is the fourth difference—killing some number within the 'too many' range would be *better* than killing some number within the 'too few' range precisely because it would make the difference between excruciating suffering and quite mild suffering, whereas killing some number within the 'too few' range would make the difference only between more excruciating and less excruciating suffering.

C15.P97 This same consideration suggests that, when it is possible to kill any number, it would be better to kill some number just outside the ‘too many’ range rather than some number just outside the ‘too few’ range. So even within the large range of options between killing too few and killing too many, some options are better or worse than others. Even if it is true that, in a pairwise comparison between adjacent options (such as killing 830 and killing 829), it is always better to kill one few rather than one more, there are other pairwise comparisons in which killing more is better than killing fewer (for example, killing 830 seems better than killing only sixty, even if sixty is outside the ‘too few’ range). Killing n more just outside the ‘too few’ range seems to matter more than killing n fewer just outside the ‘too many’ range because the suffering that would be reduced by killing some number just outside the ‘too few’ range is very great; thus reducing the suffering of the innocent victims just outside that range matters more than, or has priority over, sparing the lives of culpable Harmless Torturers, whereas the opposite may be true in comparisons between reducing suffering and sparing lives just outside the ‘too many’ range.

C15.P98 The view that emerges from these reflections is complex and in various respects indeterminate. Killing too few Harmless Torturers is always wrong provided that killing none is an option. To kill ten is to kill too few. Twenty may be too few as well, though I am not sure. There is, however, some number, perhaps fewer than fifty, the killing of whom would make a noticeable, significant difference to the suffering that the victims would experience. It would, I think, be permissible to kill that number rather than kill none. Suppose, for the purpose of illustration, that that number is forty. That might constitute a genuine threshold, at least in conditions of restricted choice. It might, for example, be permissible to kill forty but not only not better to kill thirty-nine but impermissible to kill thirty-nine, as thirty-nine would be too few, or in the zone of indeterminacy between too few and enough.

C15.P99 Beyond forty, it might be permissible to kill any number short of too many. Within this broad range between too few and too many, there is no number that it would be optimal to kill. But killing some number might be better or worse than killing some other number. This would depend on how various relevant considerations weigh against one another in the particular comparison. Among the relevant considerations are that it is worse to kill a greater number of culpable people rather than fewer, that it is more important to reduce greater suffering by some fixed amount than to reduce lesser suffering by the same amount, that a greater reduction of suffering of some rough magnitude matters more than a lesser reduction of suffering of the same rough magnitude, and so on.

C15.P100 When it is possible to kill any number of Harmless Torturers, there is, I suspect, considerable ineliminable arbitrariness in the choice among options between too few and too many. For almost *any* number one might kill within that broad range, there is some number that it would be better to kill and some

number that it would be worse to kill. (It would not be better to kill fewer than the lowest number that is clearly outside the 'too few' range and it would not be better to kill more than the lowest number outside the 'too many' range. But it might be *much* better to kill significantly more than the first of these numbers and much better to kill significantly fewer than the second.) There is no number that it would be optimal to kill, yet it would be much better, and therefore permissible, to kill any number between too few and too many than to kill too few (including none) or too many (including all). Although there are cogent objections to any choice one might make within that range, any of those choices are permissible.

C15.S7

7. The Relevance of Collaboration to Liability

C15.P101 I claimed earlier that the fact that a Harmless Torturer is *collaborating* with others in the infliction of agony on their thousand victims does *not itself* affect the wrongness of his action (though it may be correlated with other factors that do). We can see this, I think, when we recognize that the Single Torturer acts just as wrongly as any of the Harmless Torturers when he deliberately adds his tiny pain to the 999 equivalent pains that each of his thousand victims will suffer at the same time from insect bites.

C15.P102 I have also claimed that, just as collaboration makes no difference to the wrongness of a Harmless Torturer's act, so it makes no difference to his liability to defensive harming. I will conclude by briefly considering a challenge to this last claim.

C15.P103 Thus far I have considered only cases in which killing a Harmless Torturer would prevent only the effects of his own action. But we can imagine cases in which the killing of one Harmless Torturer would not only prevent him from pressing his button but would also prevent 99 others from pressing theirs, without harming them. In this case, the killing would function both eliminatively and opportunistically: it would eliminate the threat from the Harmless Torturer but also use him as a means of eliminating the threats posed by ninety-nine others.¹⁹

C15.P104 Suppose a third party has only two options: kill only this one Harmless Torturer or kill none. It seems to me, intuitively, that it would be permissible, and indeed obligatory in the absence of significant personal cost, for the third party to kill this one Harmless Torturer. Although this Harmless Torturer is not morally liable to be killed to prevent only his own contribution to the suffering of the victims, the fact that he culpably intends to contribute to the great suffering of these innocent victims makes him morally liable to be killed as a means of achieving a

¹⁹ The distinction between eliminative and opportunistic harming was introduced by Warren Quinn in 'Actions, Intentions, and Consequences: The Doctrine of Double Effect', *Philosophy & Public Affairs* 18 (1989): 334–51, at 344.

substantial reduction in their suffering. Assuming, as I have suggested, that to kill a hundred Harmless Torturers would not be to kill too few, killing any hundred of the Harmless Torturers would be proportionate and permissible. Each of the hundred would be liable to be killed as a means of achieving a reduction in every victim's suffering from a thousand to nine hundred. Assuming that this is correct, it seems that killing only one *must* be a proportionate means of achieving the same reduction. And, in the circumstances, it would also satisfy the necessity condition. This is so even though the third party would be harmfully using the one Harmless Torturer as a means. Given that the third party must choose between killing the Harmless Torturer as a means and allowing him and ninety-nine others to increase the thousand innocent victims' suffering from severe to agonizing, it seems clear that the Harmless Torturer is liable to be opportunistically used in this way.²⁰

C15.P105 But now return to the case of the Single Torturer, who knows that each of a thousand people will suffer 999 tiny pains from natural causes, such as insect bites, for eight hours. He is about to inflict one more tiny pain for eight hours on each of these victims, bringing their suffering to the same intensity as that of the victims of the Harmless Torturers. I have claimed that what he is about to do is just as wrong as what each Harmless Torturer does, and for the same reason. Yet it may seem, intuitively, that he is not liable to be killed even if killing him would not only prevent his own contribution to the suffering of the thousand innocent victims but also prevent 99 of the 999 insect bites that each victim will suffer independently of his action. How, one might ask, could it be proportionate to *kill* someone when all he would otherwise do is inflict a tiny, trivial pain on each of a thousand people, even when each of them will be in great pain quite independently of his action?

C15.P106 But this is, of course, also a description of what each Harmless Torturer does. And, for the reasons I have given above, it seems intuitively plausible to suppose that it would be proportionate, and permissible, to kill one Harmless Torturer to reduce the suffering of each of the thousand victims from 1,000 to 900. Is there a morally significant difference between an individual Harmless Torturer and the Single Torturer that can account for the difference in our intuitions about the opportunistic killing of a Harmless Torturer and the opportunistic killing of the Single Torturer?

C15.P107 One might, as I suggested earlier in Section 3, argue that the relevant difference between them is that, whereas the Single Torturer merely inflicts a tiny pain on each of a thousand victims, each Harmless Torturer is collaborating with others in the torture of a thousand victims. Because each Harmless Torturer is

²⁰ One might, alternatively, justify the killing of the one Harmless Torturer by appealing to a variant of Frances Kamm's Principle of Secondary Permissibility. See Kamm, *Intricate Ethics* (New York: Oxford University Press, 2006), 170–73.

collaborating with the others, he bears some responsibility for what they do, and this could make him liable to be killed opportunistically to prevent some of the others from pressing their buttons. The Single Torturer, by contrast, bears no responsibility for what the insects do. I argued earlier, however, that if we assume that each Harmless Torturer would have acted in the same way even in the absence of collaboration with the others, there is then no reason to suppose that any one of them is responsible for the action of any of the others, and therefore no reason to think that their collaborating affects their individual liability.

C15.P108 One might also think, as I also suggested earlier, that the collaboration of the Harmless Torturers makes each more culpable than the Single Torturer and that this affects their liability. But again we can simply stipulate that the Single Torturer and all the Harmless Torturers are culpable to the same degree. We might, for example, stipulate that the Single Torturer mistakenly believes that he is acting, not on his own, but together with 999 Harmless Torturers. Yet we might still be reluctant to accept that he is liable to be killed opportunistically as a means of preventing each of the victims in his case from suffering 99 insect bites.

C15.P109 There remain three broad options. One is to accept that it is permissible to kill a single Harmless Torturer as a means of preventing him and 99 others from pressing their buttons, that there is no relevant difference between the Single Torturer and a Harmless Torturer, and therefore that it is permissible to kill the Single Torturer as a means of reducing the suffering of the victims in that case from 1,000 to 900. Another is to reverse this reasoning and accept that it is impermissible to kill a single Harmless Torturer as a means of preventing him and ninety-nine others from pressing their buttons. The third option is to find a relevant difference between the Single Torturer and the Harmless Torturers. While I believe that we ought to accept the first of these options, there is one difference between the Harmless Torturers and the Single Torturer that could explain why killing an Harmless Torturer opportunistically is permissible but killing the Single Torturer opportunistically is not. This is that what would be prevented by killing the Harmless Torturer are harms caused by wrongdoing, while what would be prevented by killing the Single Torturer are harms resulting from natural causes. I am sceptical of the claim that the reason to prevent a harm caused by wrongdoing is stronger—or significantly stronger—than the reason to prevent an equivalent harm produced by natural causes.²¹ But many people disagree and the fact that the opportunistic killing of a Harmless Torturer would prevent a hundred acts of serious wrongdoing may well explain the difference in our intuitions about the permissibility of opportunistically killing the Harmless Torturer and the permissibility of opportunistically killing the Single Torturer.

²¹ See Jeff McMahan, 'Humanitarian Intervention, Consent, and Proportionality', in N. Ann Davis, Richard Keshen, and Jeff McMahan (eds.), *Ethics and Humanity: Themes from the Philosophy of Jonathan Glover* (New York: Oxford University Press, 2010): 44–72, at 60–2.

Bibliography

- Frowe, Helen, *Defensive Killing* (Oxford: Oxford University Press, 2014).
- Kamm, Frances, *Intricate Ethics* (New York: Oxford University Press, 2006).
- McMahan, Jeff, 'Humanitarian Intervention, Consent, and Proportionality', in N. Ann Davis, Richard Keshen, and Jeff McMahan (eds.), *Ethics and Humanity: Themes from the Philosophy of Jonathan Glover* (New York: Oxford University Press, 2010): 44–72.
- McMahan, Jeff, Review of Frowe, *Defensive Killing*, in *Ethics* 126 (2016): 825–31.
- McMahan, Jeff, 'The Limits of Self-Defense', in Christian Coons and Michael Weber (eds.), *The Ethics of Self-Defense* (New York: Oxford University Press, 2016): 185–210.
- McMahan, Jeff, 'Proportionate Defense' (revised and expanded version), in Jens Ohlin, Larry May, and Claire Finkelstein (eds.), *Weighing Lives in War* (Oxford: Oxford University Press, 2017): 131–54.
- Parfit, Derek, *Reasons and Persons* (Oxford: Clarendon Press, 1984).
- Parfit, Derek, *Reasons and Persons* (Oxford: Clarendon Press, 1987 repr.).
- Pettit, Philip, 'Responsibility Incorporated', *Ethics* 117 (2007): 171–201.
- Pummer, Theron, 'Sorites on What Matters', in Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan (eds.), *Ethics and Existence: The Legacy of Derek Parfit* (Oxford: Oxford University Press, forthcoming).
- Quinn, Warren, 'Actions, Intentions, and Consequences: The Doctrine of Double Effect', *Philosophy & Public Affairs* 18 (1989): 334–51.
- Temkin, Larry, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning* (New York: Oxford University Press, 2012).