



The concept of responsibility in the ethics of self-defense and war

Carolina Sartorio¹

Accepted: 18 January 2021 / Published online: 3 February 2021

© The Author(s), under exclusive licence to Springer Nature B.V. part of Springer Nature 2021

Abstract The focus of this paper is an influential family of views in the ethics of self-defense and war: views that ground the agent’s liability to be attacked in self-defense in the agent’s moral responsibility for the threat posed (“Responsibility Views”). I critically examine the concept of responsibility employed by such views, by looking at potential connections with the contemporary literature on moral responsibility. I start by uncovering some of the key assumptions that Responsibility Views make about the relevant concept of responsibility, and by scrutinizing those assumptions under the lens of more general theorizing about responsibility. I identify an important conflict that arises at that point. The problem is that the concept presupposed by Responsibility Views is in tension with the standard way of understanding the connection between the neutral and non-neutral forms of moral responsibility. I draw attention to a particular strategy that could be used to address this challenge, but I also identify some important obstacles that stand in the way.

Keyword moral responsibility · defensive harm · self-defense · ethics of war · blameworthiness · liability

1 Introduction

The focus of this paper is an influential family of views in the ethics of self-defense and war: views that ground the agent’s liability to be attacked in self-defense in the agent’s *moral responsibility* for the threat posed. I’ll refer to them as “Responsibility Views.” My main aim is to critically examine the concept of responsibility employed by such views, by looking at potential connections with the contemporary

✉ Carolina Sartorio
sartorio@arizona.edu

¹ The University of Arizona, Tucson, AZ, USA

literature on moral responsibility (which these debates surprisingly tend to stay away from, almost completely). I'll proceed by uncovering some of the key assumptions that Responsibility Views make about the relevant concept of responsibility, and by scrutinizing those assumptions under the lens of more general theorizing about responsibility. I'm particularly interested in examining how the literatures on self-defense/war and responsibility could potentially be brought together in this case, as well as in drawing attention to the challenges that arise in doing so.

The structure of the paper is the following. In Sect. 2, I outline the main claims made by Responsibility Views, and the motivations behind them. In Sects. 3 and 4, I discuss the concept of responsibility presupposed by those views, and the connections with the general literature on responsibility. As we will see, there is an important conflict that arises at that point, and so in Sect. 5 I explore a possible way to resolve it, as well as some challenges that remain. I end with some concluding remarks in Sect. 6.

2 Responsibility views

A recent trend in the ethics of just war, and the ethics of defensive harm that underlies it, is an appeal to *agential* concepts. According to those who embrace such agency-based views, the right account of self-defense must draw attention to the agency—in particular, the *moral* agency—of those who will be harmed by defensive force. For, some think, this is the best way to support plausible restrictions on what we can permissibly do in self-defense. In particular, it is necessary to distinguish cases of permissible self-defense from impermissible cases of harming mere “bystanders.”¹

The most natural way to appeal to moral agency in an account of permissible defense is to embrace a *culpability* view: a view according to which, roughly, the culpability or blameworthiness of an attacker is what grounds the permissibility of defensive harm against her (McMahan 1994; Rodin 2002; Ferzan 2005; Alexander and Ferzan 2009). But a common objection to these views is that this is too restrictive. For culpability views entail that defensive force is impermissible in, for example, cases of this kind:

Conscientious Driver: Threatener owns a car, which she carefully maintains, following all of the manufacturer's instructions. You are a pedestrian, walking down the sidewalk. Threatener drives to the grocery store for some supplies and, despite her best efforts, the car malfunctions and skids towards you, who will be killed by the impact. You have a weapon, which you can use to vaporize Threatener's car. This is the only way in which you can defend

¹ See the overview of this literature in Doggett (2011). Doggett identifies the appeal to agency (the agency of those whom you would harm in self-defense) as one of the most important developments in the contemporary self-defense literature.

yourself, and it will kill Threatener. (Adapted from McMahan 2009: 165, and Hosein 2017.)

The conscientious driver is not at all culpable for the threat to your life; still, it seems clearly permissible for you to kill her in self-defense. And in contexts of war, too, when soldiers who are fighting an unjust war do so non-culpably, it seems nonetheless permissible for soldiers on the just side to kill them in self-defense. But, again, culpability views would entail that the killing of non-culpable soldiers is impermissible.

Responsibility Views attempt to deal with this problem by providing a more permissive account, but one that is still based on the agent's moral agency. According to these views, the extent of someone's liability to defensive harm, and thus, in many cases, the permissibility of defensively harming her, is determined by the agent's responsibility for the threat posed,² not her culpability. Those who are liable to harm lack rights against being harmed. Thus, while it is not only liable people who may be permissibly harmed, it is much easier to justify harming a liable person than a non-liable person.

Importantly, the relevant concept of responsibility is (like culpability) also supposed to be *moral*. That is to say, it's supposed to concern the exercise of certain capacities that only moral agents have, and the assessment that an agent is responsible for a threat is supposed to tell us something about the agent *qua* moral agent. But the relevant concept of responsibility is still significantly broader than culpability in that it is possible to be responsible for posing a threat without in any way being culpable, as Conscientious Driver is supposed to illustrate.

According to Responsibility Views, Threatener in Conscientious Driver is morally responsible (without being culpable or blameworthy) for the threat that she poses to you because she voluntarily engaged in an activity—driving—that is known to be associated with an increased risk of endangering lives, even if the risk is very small. Given that she did this voluntarily, and the risk (unluckily) eventuated, she is morally responsible for the threat to your life (without being culpable). This is what makes it permissible for you to kill her in self-defense. Similarly, when soldiers fight a war, there is always some risk that they will threaten unjustified harm. To the extent that they can be in a position to realize this, they can be morally responsible for the unjust threats that they pose (without being culpable). Their responsibility can render them liable to defensive harm. Again, it is important, on this view, that we understand the relevant concept of responsibility as moral. Although the agent is not culpable for posing the threat, the thought is that freely and voluntarily taking “moral risks” is sufficient for liability, because it's sufficient for bearing at least some moral responsibility for the manifestation of the risk, when the risk manifests in the form of an actual threat to others.

The most developed and influential version of the Responsibility View is McMahan (2009) (but see also Otsuka 1994, 2016, and Hosein 2017). McMahan's view is quite complex, in that it combines the claim that responsibility is what grounds liability with another important idea: namely, that both responsibility and the degree of harm to which one can be liable come in degrees. McMahan also holds that

² Or, at least, for the *appearance* of a threat. For simplicity's sake, here I'll focus on actual-threat cases.

increasing degrees of responsibility in turn result in liability to increased harm. This gives rise to more fine-grained discriminations of liability. Thus, on McMahan's view, although both culpable and merely responsible threateners can be liable, culpable threateners are liable to larger degrees of harm than merely responsible threateners because they are *more* morally responsible for the threat that they pose. To say that they are liable to larger degrees of harm means that culpable threateners may be subjected to a greater amount of force than merely responsible threateners. For example, if you can prevent a person from running over your foot only by killing her, it might be permissible to do this if she threatens intentionally and culpably, but impermissible if she threatens unintentionally and non-culpably. So, on McMahan's view, the agent's responsibility matters not merely for determining whether an agent is liable, but also for determining the degree of harm to which she is liable. In fact, degrees of responsibility range from the highest degrees of culpability (intentional, unjustified, and unexcused threatening behaviors), through some intermediate cases of limited culpability (threatening behaviors that are partially excused in different ways), all the way to the lowest degrees of responsibility (fully excused, non-culpable threatening behaviors, for which the agent is still morally responsible).³

As McMahan recognizes, although his proposal is more permissive than the culpability view, it is also quite restrictive. For it entails that it's impermissible to defend yourself in cases where the person who is threatening you is *not at all morally responsible for the threat*, as in the following cases:

Falling Person: A villain has thrown an innocent person, Threatener, from a tall building and her falling body will crush you to death (while she will survive the fall herself), unless you vaporize her with your ray gun. (Adapted from Thomson 1991 and Otsuka 1994; originally inspired by Nozick 1974: 34–5.)

Cell Phone: A villain has rigged Threatener's cell phone, so that if she calls anyone a bomb strapped to you will detonate. She is completely unaware of this and has no reason to suspect it. She is now dialing someone's number. The only way in which you can save yourself is to kill her using your weapon. (From Hosein 2017; adapted from McMahan 2009: 165.)

In Falling Person, the perfectly innocent Threatener is not exercising her agency at all when she threatens you, so she is clearly not morally responsible for the threat that she poses to you. She is, of course, *causally* responsible for the threat. But recall that Responsibility Views focus on moral responsibility, not just causal responsibility: they claim that the agent is responsible for the threat not only in the sense that she causally contributes to it, but also in the sense that this is somehow reflective of her as a moral agent. Still, many think it's permissible for you to kill the falling person in self-defense.

³ See McMahan 2009, chapter 4. In light of this, Lazar (2010) calls culpability cases "maximal responsibility" cases, and non-culpable responsibility cases "minimal responsibility" cases (sometimes he uses "agent-responsibility" instead of "minimal responsibility"; see Lazar 2009, 2010). I'll also use "plain" responsibility to refer to this kind of responsibility. Although this scalar model of responsibility and culpability has been highly influential, some would disagree with its central tenet (see, e.g., Tadros 2018).

In turn, in Cell Phone, Threatener is exercising her agency when she is starting to make a call, and, according to McMahan, she is responsible for starting to make a call. But, McMahan also thinks, she isn't responsible for *the threat that the call poses to you*, since she is completely unaware of this. Moreover, one is supposed to assume that, unlike the conscientious driver, no reasonable person would associate making a call with an increased risk of causing harm to others.⁴ So, again, Responsibility Views entail that it's impermissible to kill the cell phone user, when it seems that it's not. Although McMahan recognizes that these implications of his view are somewhat counterintuitive, he bites the bullet and goes on to give arguments for their truth (as does Otsuka in his 1994).⁵

Most critical discussions of Responsibility Views focus on the implications of those views and on the plausibility of the arguments that are used to support those implications.⁶ In contrast, my way of proceeding will be to look, more fundamentally, into the concept of responsibility that is presupposed by these views and the role that it plays in them. Given that what is at stake is supposed to be a form of moral responsibility (albeit a potentially neutral one that can come apart from blameworthiness or culpability), I will start by looking at the relevant parts of the general responsibility literature. I turn to this in the next section.

3 Neutral/non-neutral forms of responsibility, and the standard view of their relation

For the most part, the general literature on moral responsibility focuses on a *non-neutral* conception of responsibility: typically, blameworthiness, although sometimes praiseworthiness is discussed alongside blameworthiness.⁷ This is not surprising, since at least a big part of the interest in moral responsibility stems

⁴ What McMahan seems to have in mind is that it's common knowledge that driving a car at a certain speed increases the probability that one will pose a threat to someone, even it is by some small magnitude, whereas making a phone call doesn't. I'll go along with these assumptions here.

⁵ On the other hand, Hosein (2017) aims to develop a version of the responsibility view that avoids these counterintuitive implications. On his view, non-responsible threats can be liable, but their lack of responsibility results in their being liable to lesser harms than other threats.

⁶ See, e.g., Lazar (2009, 2010), and the replies to McMahan by Haque, Wallerstein, Ferzan, and Tadros in Robinson et al. (2009, chapter 18). Kaufman's reply to McMahan (Kaufman 2009) is an exception in that he suggests that more would have to be said about the relevant concept of responsibility (and, in particular, about how it differs from culpability) before one can give a full assessment of McMahan's responsibility view.

⁷ The focus is usually on moral responsibility as a form of *accountability*—that is to say, a notion of responsibility in light of which we can be held to account for what we do. Sometimes this is distinguished from *attributability*—a notion of responsibility in light of which our behavior can be attributed to us in that it reflects certain morally significant features of our character, such as our virtues and vices. (For a discussion of this distinction, see Watson 1996.) I'll set attributability aside here, since it's arguably not the form of responsibility at stake in these debates. For example, the conscientious driver is not responsible for her threatening behavior in a sense that is revelatory of any moral vices on her part. Still, the thought is that she can be held to account given that she acted with an understanding of the risk involved.

from considerations having to do with the problem of free will and the related question of whether it's at all appropriate to blame (or praise) human beings for what they do. In contrast, as we have seen, the views we're interested in here appeal to a potentially neutral form of responsibility—in particular, one that cannot be identified with culpability or blameworthiness. It's potentially neutral in the sense that it needn't involve the judgment that the agent is blameworthy or praiseworthy, to *any* degree, but it's still supposed to be a form of moral responsibility. So, one fundamental question that immediately arises concerns this basic assumption: *Is there* such a concept of responsibility, and one that can play the role that responsibility theorists attribute to it?

Although, again, a neutral form of responsibility is rarely the focus of moral responsibility debates, it is sometimes acknowledged as a legitimate concept (and occasionally worth talking about) but rarely discussed in much detail. Fischer has a brief discussion of it in his work on free will and responsibility (partly in collaboration with Ravizza). His view of responsibility is inspired by Strawson's conception of the reactive attitudes (Strawson 1962). Strawson famously suggested that being morally responsible is being an apt target of attitudes such as praise and blame, gratitude and resentment, etc. Fischer assumes a view of this kind, but suggests that there is an important difference between being an *apt* (potential) target for reactive attitudes and being an *actual* target (Fischer 2004; see also Fischer and Ravizza 1998: 6–8). This results in a wedge between responsibility and blameworthiness or praiseworthiness. As Fischer sees it, responsibility is merely the gateway to blameworthiness or praiseworthiness: besides being morally responsible, other conditions need to obtain in order for blame or praise to be warranted. In other words, blameworthiness and praiseworthiness entail responsibility, but not the other way around.

Now, it's important to note that this isn't just supposed to be the claim that sometimes it's not appropriate for us to blame or praise others for what they do, even if they are morally responsible for their behavior. In the literature on responsibility and blame, it is common to distinguish between agents being *blameworthy* for what they do, on the one hand, and *blame being a fitting response* to that behavior by someone on a particular occasion, on the other hand (see, e.g., Coates and Tognazzini 2012a, b). Even when people are blameworthy for what they do, it may not be appropriate for me, a potential blamer, to blame them for what they do. This could be for a variety of reasons. For example, I may not have the relevant moral authority to blame them (perhaps I share some of the blame for what they did, or perhaps I would've done exactly the same thing if I'd been in their shoes). For whatever reason, blame may not be a fitting attitude for me to have towards the transgressor, in light of my special circumstances. Given this, it's not difficult to see how responsibility and specific instances of blame can come apart. But what we're interested in here is not specific instances of blame but, rather, *blameworthiness*: that is, the more basic issue of blame being at all warranted, or agents being deserving of blame, given what they did.

So, what motivation could there be for wanting to distinguish responsibility from blameworthiness in this kind of way? What most obviously comes to mind is the fact that sometimes we exercise our free agency in the form of behaviors that are

morally neutral, and that result in consequences that are themselves morally neutral.⁸ On reflection, most of what we do in our daily lives seems to be morally neutral in this way. It might make sense to think, for example, that we are morally responsible for spending the weekend reading a book. Nothing blameworthy or praiseworthy about it (in most cases), but still a behavior that we can take responsibility for—at least if we did it voluntarily, nobody forced us to do it, etc.

One may add the following in support of the idea that we can be responsible for these morally neutral behaviors: in many of these cases, while acting in morally neutral ways, we're still responding to the fact that we're not harming anybody or violating anybody's rights in so acting. In other words, we are aware of the fact that the behaviors are morally neutral, and this is partly why we engage in those behaviors to begin with (otherwise, if we had suspected that we were likely to cause harm to somebody or violate somebody's rights, we wouldn't have acted in those ways, or we would've given it more thought before doing so). Thus, we're still exercising our agency in a way that reflects on us as moral agents, and can be morally responsible for what we do, in at least this minimal sense.⁹

McKenna (2012: 15) discusses other behaviors that are not morally neutral because they are in fact morally *required*, and we tend to be fully aware that they are required; however, we are still not praiseworthy for them because they only meet the basic standards of moral decency. For example, presumably you are morally responsible for not murdering your neighbors every time you see them, but you are not praiseworthy for that (and of course you are not blameworthy for that either). The fact that praiseworthiness seems to require more than meeting the minimal standards of moral decency results in this category of behaviors for which we are morally responsible without being praiseworthy (or blameworthy) for them. Again, a lot of what we do (or fail to do) in our ordinary lives seems to be of this kind: we interact with many people on a regular basis and we are morally decent to them. Thus, we are morally responsible for all of that, without, arguably, being blameworthy or praiseworthy for any of that.

All of this suggests that there is, in fact, a quite general *neutral* concept of responsibility. As a result, it might be tempting to identify this concept with the one used by responsibility theorists in the ethics of self-defense and war. But at this point an important problem arises. For, at the same time that the previous observations can help us see how responsibility can come apart from blameworthiness and praiseworthiness, they can also help us think about the *relation* that exists between the neutral and non-neutral forms of responsibility. And, as we will see, the most natural account of that relation clashes with Responsibility Views. In the remainder of this section, I'll explain what that account is. Then, in the following section, I'll explain how the conflict with Responsibility Views arises.

⁸ For example, Fischer and Ravizza (1998: 8, n. 11) mention the example of raising your hand on one occasion where this doesn't have any moral implications whatsoever.

⁹ In other words, the thought is that we can be responsible for what we do to the extent that, in acting in any of those morally neutral ways, we're still responding to moral reasons, or to the *absence* thereof (for a discussion of the role played by absences of reasons in our moral responsibility, see Sartorio 2016).

What is the connection between the neutral and non-neutral forms of moral responsibility? Surely, there must be *some* connection. And it seems very natural to assume, on the basis of the above examples, that it's simply this: blameworthiness and praiseworthiness are the forms of responsibility at stake whenever we are dealing with behaviors and outcomes that are morally "loaded" in the relevant sense.

Call this the *Standard View* of the relation between the neutral and non-neutral forms of responsibility. Here is a possible formulation of such a view:

Standard View:

Blameworthiness = Moral responsibility for wrongful acts and for bad outcomes resulting from those acts.

Praiseworthiness = Moral responsibility for morally right acts (or supererogatory acts, or acts that are good beyond a certain threshold) and for good outcomes (or for outcomes that are good beyond a certain threshold) resulting from those acts.

For example, assuming an objectivist account of wrongness (i.e. one that is "fact-based" instead of "evidence-based"), the Standard View says that we are blameworthy for what we do when we are morally responsible for what we do, and what we do is (objectively) wrong. Since most views of blameworthiness require objective wrongness, in what follows I'll adopt this assumption.¹⁰

Notice that one of the consequences of the Standard View is that blameworthiness and praiseworthiness entail responsibility, but not the other way around. For all the view says, we could be morally responsible without being blameworthy or praiseworthy. In particular, we could be responsible for morally neutral acts, or for acts that are just morally decent.¹¹

The Standard View is initially very plausible. Most theorists interested in the general concept of responsibility seem to embrace it, at least implicitly. It is, after all, the simplest and most natural view about the relation between the neutral and

¹⁰ This is sometimes called "the objective view" of blameworthiness (a similar claim is made about praiseworthiness). Some proponents of the objective view include Copp (1997), Smith (1983), Wallace (1994), and Widerker (1991). McKenna (2012: 14) identifies it as the most commonly accepted view. Even those who reject it (proponents of "subjective views" of blameworthiness) tend to see themselves as arguing against tradition, and refer to the target of their attack as a widespread, initially appealing, and highly intuitive view (see, e.g., Capes 2012; Haji 1998: 141, and Khoury 2011). To keep things as simple as possible, I won't be concerned with subjective views in this paper. But it's important to note that at least some subjective views have trouble accommodating the conscientious driver case. For example, Zimmerman's view, which allows for neutral responsibility judgments, entails that the conscientious driver is culpable for posing a threat, at least to some degree, given her *belief* that she was taking a moral risk by driving the car (see Zimmerman 1988: 49 and 61–62). On the other hand, other subjective views that ground blameworthiness in expressions of an objectionable quality of will or the negative reactive attitudes they elicit (e.g., Capes 2012) entail that the conscientious driver isn't blameworthy, but it's harder to motivate, within the framework of such views, a *neutral* concept of responsibility that could distinguish between the conscientious driver and the cell phone user.

¹¹ We could also be morally responsible (without being blameworthy) for bad outcomes that are the result of behaviors that aren't wrong. For example, if you are in a trolley problem situation and you switch the trolley that was going to kill five towards one, you are arguably responsible for the one person's death without being blameworthy for it. For discussion of this issue, see Sartorio Forthcoming.

non-neutral concepts. As a result, it seems to be the “default” view about that relation, and we should embrace it unless we can think of good reasons to believe otherwise.

Now, when I said that most theorists interested in the general concept of responsibility accept the Standard View, I meant to make room for one important exception: Fischer’s own view (I mention it here because it’ll play a central role later, when we revisit Responsibility Views). Fischer (2004) argues that agents can be morally responsible for wrongful acts without being blameworthy for them. He thus argues for a wider wedge between responsibility and blameworthiness than the Standard View allows for. On Fischer’s view, conditions having to do with how our characters were formed or how we came to acquire our values can be relevant to whether we are blameworthy, even if the minimal threshold of control and responsibility has been met. He has in mind pretty unusual and special cases where agents have been manipulated by an external source, but in a way that leaves their major capacities for agency and reasons-responsiveness intact. As a result, according to most compatibilist accounts of responsibility, despite having been manipulated, these agents are still responsible for their wrongful behaviors. This is part of Fischer’s reply to Pereboom (2001), where Pereboom argues that causal determinism and responsibility are incompatible on the basis of such manipulation cases.¹² I won’t go into the details of this debate here. But an important part of Fischer’s reply is to concede that those manipulated agents are not blameworthy for their wrongful behaviors, but to argue that they can still be responsible for them.

Now, of course, this raises the question of *what else* could be required for blameworthiness, besides being responsible for a wrongful act. Again, given how initially plausible the Standard View is, if we are going to abandon it, we need good reason to do so. In addition, we need to have some idea of what could work as a replacement—and the less the replacement looks like an outright rejection and the more it looks like a revision, arguably the better.

Unfortunately, Fischer doesn’t have much else to say about this.¹³ But note that one possibility would be to suggest that control over actions comes in degrees. Assuming there are degrees of control, Fischer could say that agents who have been manipulated in the way Pereboom was imagining are less in control of their acts than other agents who are not manipulated. On that basis, Fischer could argue that responsibility also comes in degrees as a result, and that what needs to be added to the Standard View is the idea that blameworthiness requires a certain *threshold of responsibility*, a threshold that some responsible agents may not reach. Perhaps Pereboom’s manipulated agents don’t reach that higher responsibility threshold (because they don’t reach the relevant control threshold). As a result, they are

¹² Pereboom argues that those manipulated agents are not responsible, and, on that basis, that causally determined agents are also not responsible.

¹³ One thing that he *couldn’t* plausibly say is that the manipulated agents are not blameworthy simply because it would be inappropriate for most of us (non-manipulated agents) to blame them. For recall that we’re assuming that there is an important distinction between agents being blameworthy and others having the relevant standing to blame them. For a discussion on this distinction as it bears on the responsibility of manipulated agents, see King (2015).

responsible to some degree without being blameworthy. In principle, it looks like this could result in a relatively minor modification of the Standard View, one that retains most of the initial plausibility of the original view.

Now, I should note that this isn't likely to be Fischer's own view, since his official account of responsibility (in Fischer and Ravizza 1998, as well as in other work on his own) is not scalar but on/off. Never mind; it seems to be a view at least worth considering. Plus, recently there has been a growing literature attempting to analyze responsibility as a scalar concept, or at least taking some initial steps in that direction.¹⁴ Some of these accounts actually build on Fischer and Ravizza's own view of control, so they could be seen as a natural extension of their view.

However, more would need to be said to motivate a proposal of this kind. Imagine that there are, in fact, degrees of control. Why think that those degrees work in the way suggested above—that is, that they result in degrees of responsibility, and that blameworthy agents are only those agents who are responsible *beyond a certain threshold*? Instead, it seems much more natural to suggest that, if there are degrees of control, they result, more simply, in degrees of *blameworthiness*, whenever agents are responsible for their wrongful behavior. This is a very natural way of understanding the role that degrees of control could potentially play in a theory of responsibility. And, it is compatible with the Standard View. So, again, given that the Standard View is the default view of the relation between responsibility and blameworthiness, we would need a stronger reason to depart from it: simply noting that there are degrees of control isn't enough.

To sum up, in this section I have reviewed some central work that has been done in the general responsibility literature on a potentially neutral concept of moral responsibility. We have seen that it is plausible to think that there is an intelligible concept of this kind. We have also examined the question about the relation between the neutral and non-neutral concepts. I suggested that the most natural way to understand their relation is by means of what I called “the Standard View.” I then identified Fischer as somebody who would reject such a view, but I argued that Fischer's position isn't sufficiently well motivated as it stands, and that more would have to be done to motivate a departure from the Standard View.

In the next section, I return to Responsibility Views in the ethics of war and self-defense. I show that, like Fischer's view, Responsibility Views also need to assume a wedge between responsibility and blameworthiness, although this time it's a different kind of wedge. As a result, these views must also reject the Standard View, but they must do so for different reasons.

¹⁴ See, e.g., Björnson (2017), Coates and Swenson (2013), Nelkin (2016), Guerrero (2017), Tierney (2019), Usher (2020), and Kaiserman (Forthcoming). An important challenge that arises for such views is that, if one wants to offer a comprehensive view of responsibility as a scalar notion, there are different dimensions that bear on an agent's responsibility and that would have to be weighed against each other: how much control the agent had, how much the agent knew, how good/bad the agent's intention was, etc.

4 The conflict between responsibility views and the standard view

First of all, when thinking about cases of defensive force, it's important to keep track of another distinction: the distinction between an agent's responsibility for a *behavior* leading to an outcome and her responsibility for the *outcome* itself. It is commonly acknowledged that, when we are responsible for outcomes, our responsibility is derived from our responsibility for the behaviors issuing in those outcomes. That is to say, outcomes are not things that we can be "basically" responsible for, but things that we get to be responsible for by virtue of, more fundamentally, being responsible for other things (behaviors, choices, etc.). This is relevant for our purposes here because our focus in the context of the ethics of permissible defense is *outcome-responsibility*, not basic responsibility. The main question we're interested in is, again, not whether agents are responsible for engaging in certain behaviors, but, instead, whether agents are responsible *for the threat that those behaviors pose*. This is responsibility for an outcome—namely, the threat posed by a behavior—not responsibility for a behavior.

As suggested above (in Sect. 2), this distinction plays an important role in a case like Cell Phone. This is worth repeating here. In that case, by assumption, the agent is supposed to be responsible for the behavior (making a call) but not for one of its outcomes: the threat posed by that behavior. And, again, according to Responsibility Views, it's the responsibility for the *threat*, not the responsibility for the behavior itself, that is directly relevant to her liability. This is why the cell phone user is assumed to be non-liable: because she's not responsible for the threat she poses.

So, what we're interested in here are questions of this kind: When one behaves in a way that results in a threat being posed, what is the relation between being responsible *for the threat posed* and being blameworthy for it? What else is required, in order to be blameworthy for the threat, besides being morally responsible for it? Only if there is such a distinction between being responsible for an outcome and being blameworthy for an outcome, can we sustain the claim that the conscientious driver is responsible, but not blameworthy, for posing a threat.

As we have seen, according to the Standard View, your being blameworthy for an outcome amounts to such an outcome's being bad and its resulting from some wrongful behavior of yours for which you are responsible. Now, the threat that the driver poses is clearly a bad thing, and it results from some (objectively) wrongful behavior of hers for which, according to proponents of Responsibility Views, she is responsible.¹⁵ So, on the basis of these assumptions, the Standard View would entail that she is also blameworthy for it. But she *isn't* blameworthy for the threat. This means that responsibility theorists must reject the Standard View.¹⁶

¹⁵ The behavior is wrongful but not, of course, blameworthy (it is wrongful for the driver to act in a way that poses a threat, even if she's completely blameless for that behavior). Recall that we're assuming an objectivist account of wrongness (see Sect. 3 above).

¹⁶ At this point, responsibility theorists could think of proposing an interpretation of the Standard View according to which blameworthiness for an outcome doesn't just require responsibility for the wrongful behavior but *blameworthiness* for the wrongful behavior. On that basis, they could note that it doesn't follow from the Standard View that the driver is blameworthy for the threat, because she's not

What would the Standard View say about the conscientious driver? Basically, this: although the conscientious driver acted wrongly when she drove in a way that posed a deadly threat to someone, she isn't morally responsible for her wrongful behavior, or for the threat posed by her behavior, given that she did it unknowingly and non-culpably. In the general literature on moral responsibility, it is common to distinguish two components of, or conditions for, responsibility: one has to do with our control and the other with our epistemic state. Roughly, this means that being responsible for a wrongful behavior requires both the right kind of control over what you're doing and the right kind of awareness of what you're doing and of the moral significance of what you're doing (or, at least, it requires that you *should* have been aware of such facts¹⁷). And, although the conscientious driver was certainly in control of at least some central aspects of her behavior, she was not aware of the threat that such a behavior would pose, and reasonably so. As a result, she is not morally responsible for her wrongful behavior, or for the threat that resulted from that wrongful behavior.¹⁸

Notice that it doesn't follow from this that the conscientious driver isn't morally responsible for anything at all. She can still be responsible for certain *morally neutral* aspects of her behavior—for example, for certain bodily movements that she made that did not contribute to her posing a threat to anybody, such as her holding the steering wheel in a certain way, her going at a certain speed, etc. But, according to the Standard View, she cannot be morally responsible for the *wrongful* aspects of her behavior, or for the threat that they resulted in, given that she is not blameworthy for these.

According to the Standard View, then, in some important respects the conscientious driver is just like the cell phone user. For the cell phone user, too, wrongfully posed a threat to someone without being morally responsible for it. And the reason she is not responsible for posing that threat is, again, that she was not aware that she would pose such a threat, and reasonably so. Again, although she was in control of certain aspects of her behavior, this is not enough for her to be responsible for the wrongful aspects of her behavior, or for the threat posed by them. As a result, although the conscientious driver and the cell phone user might still come apart in other ways (more on this below), they are on a par with respect to their *moral responsibility*: neither is morally responsible for their wrongful behavior, or for the threat that such a wrongful behavior posed.

Footnote 16 continued

blameworthy for that wrongful behavior. But this won't work as a response, because the Standard View would in fact entail that she *is* blameworthy for the wrongful behavior (given that we're assuming that she is responsible for it). So, this move would just push the problem back instead of eliminating it.

¹⁷ This is how cases of negligence are typically accommodated within a theory of responsibility. (For further discussion, see, e.g., the contributions in Robichaud and Wieland (2017).) By assumption, the conscientious driver wasn't acting negligently when she posed the threat.

¹⁸ In contrast, note that the cases of neutral responsibility we examined before (in Sect. 3) are different in that in those cases the agents arguably meet the epistemic condition for responsibility. For example, typically, when acting in morally neutral ways, you're reasonably aware that you're not harming anybody or violating anybody's rights by acting in those ways.

At least, this is what the Standard View would say. But perhaps the Standard View should be given up? In the next section, I explore a way in which proponents of Responsibility Views could argue for this, in an attempt to resolve the problem.

5 A potential resolution of the conflict, and some remaining challenges

I see three main ways in which proponents of Responsibility Views could proceed at this point.

First, they could say that they never intended for their concept of responsibility to track the more general concept of responsibility—that is, the concept with which theorists working on the free will problem, for example, are concerned. In that case, it shouldn't be seen as a problem for Responsibility Views that they clash with the Standard View, if Responsibility Views work with a different concept altogether.

Second, proponents of the Responsibility View could retract their claim that an agent's liability is grounded in her moral responsibility for the threat posed. Note that, even if the Standard View were right, there could still be a difference between the conscientious driver and the cell phone user, one that concerns the agents' understanding of the risk involved.¹⁹ In that case, even if the conscientious driver weren't, as a moral agent, *responsible for the threat* in any meaningful way, her liability could still be grounded in a different fact: the fact that she understood the increased risk involved. In other words, this solution amounts to claiming that the driver's liability is not at all grounded in her moral responsibility, but in the relevant features of her epistemic state.

Both of these moves are quite radical. The first results in an implausible disconnection between Responsibility Views and the concepts used in the more general responsibility debates. I find it highly implausible to think that the two literatures are so disconnected that they're simply concerned with different concepts. The second, in turn, results in a significant departure from the initial motivation for Responsibility Views. For recall that these views are typically motivated by the wish to improve on the more extreme culpability-based views *while preserving a central tenet of those views*: the claim that liability is grounded in a form of moral responsibility (see Sect. 2 above). But opting for this solution to the conflict amounts to giving up that central claim.

For these reasons, I won't be pursuing these options any further. Instead, in the remainder of this section I'll explore a different way in which proponents of Responsibility Views could hope to address the challenge. This third route consists in identifying a motivation for revising the Standard View in such a way that the revised version ends up being consistent with Responsibility Views. I'll discuss a specific strategy for doing this, which draws on the earlier discussion of Fischer's view in Sect. 3 above.

¹⁹ Recall that we're assuming, with McMahan, that it's reasonable to believe that driving a car increases the risk of posing a threat to others, even if it's by some small magnitude, whereas it's not reasonable to believe this about making a phone call, and that the agents were aware of all of that. See note 4 above.

Recall that responsibility is thought to have two components, which are captured by a control condition and an epistemic condition. The control condition states that, in order for us to be responsible, we must be in control of what we do, and the epistemic condition states that we must be aware of the moral significance of what we do. And recall that, during the discussion of Fischer's view in Sect. 3, I noted that one way in which one could argue against the Standard View is by appealing to a graded notion of *control*, and to the idea that blameworthiness requires a higher threshold of control, and thus of moral responsibility, than plain responsibility. If one could motivate this idea, then one could argue that being blameworthy amounts to more than just being responsible for a wrongful act: it amounts to being responsible for a wrongful act that is under one's control *beyond a certain threshold*.

What I want to suggest now is that proponents of Responsibility Views could argue against the Standard View in a similar way, by focusing on the *epistemic condition* of responsibility instead. Although, as far as I know, nobody has suggested something like this in the general literature on moral responsibility, it is a view that seems at least worth considering.

How, exactly, could proponents of Responsibility Views argue for this? Following McMahan's scalar model of responsibility/culpability described in Sect. 2, they could claim that the relevant kind of *awareness*—that is, one's understanding of the moral significance of the situation—also comes in degrees, and that blameworthiness requires a higher “awareness threshold” than plain responsibility. In particular, they could claim that the conscientious driver doesn't meet this awareness threshold for blameworthiness, despite meeting the threshold for responsibility, because she is unaware of certain key facts about the moral significance of what she is doing. In particular, proponents of the Responsibility View could argue, the conscientious driver is unaware of the facts that make it the case that, on this particular occasion, she'll end up posing a threat. As a result, the conscientious driver is *less aware* of the moral significance of what she is doing than somebody who knew all the relevant facts (a “fully aware” threatening driver, say) because she is aware of fewer of those morally significant facts. This could be used to explain why she doesn't reach the threshold for blameworthiness.

If this proposal could be sufficiently motivated, proponents of the Responsibility View could, on this basis, reject the Standard View and propose a revision. They could suggest that being blameworthy amounts to more than just being responsible for a wrongful act: it amounts to being responsible for a wrongful act when our awareness of the moral significance of what we're doing *surpasses a certain threshold* (in the sense explained above). Although such a “threshold version” of the Standard view is a departure from the original version, they could argue that it's sufficiently close to it to remain a plausible contender.

Again, however, more would have to be said to motivate this view. I'll conclude by identifying what I see as the main challenges that arise at this point.

First, one would have to do more to motivate the idea that what is doing the work here is the relevant *degrees of awareness*. The proposal under consideration is that, in order for the conscientious driver to reach the “awareness threshold” that would result in her being culpable, she would have to have been *more aware* of the moral significance of what she was doing, by, in particular, being aware of more of the

relevant facts—including, notably, the fact that she would end up posing a threat on this particular occasion. However, merely noting that, if she had been aware of that fact (i.e. if she had been a fully aware driver), then she would have been culpable for posing a threat isn't sufficient to motivate the idea that the difference between the fully aware driver and the conscientious driver is a difference in degree. Clearly, there is an important difference in *what* each of them knew. But it's not obvious that this results in a difference in the relevant degrees of awareness. For all that has been said, it's just a difference *in kind*: it's the difference between knowing that you'll actually pose a threat and not knowing that.²⁰

Second, the same type of problem that arises for Fischer's proposal arises here too. Imagine that there are degrees of awareness, and that they work in roughly the way described above (i.e. the fully aware driver is more aware of the moral significance of what she's doing than the conscientious driver). Still, a major obstacle remains. For, even then, one would have to do more to motivate the idea that those degrees of awareness plausibly result in a difference between being merely responsible and being culpable.

Basically, the problem is this: Why think that degrees of awareness would work in this way, if they existed? Perhaps there are such degrees of awareness, but they simply result in degrees of *blameworthiness* (whenever an agent is in fact responsible for their wrongful behavior). This seems to be a much more natural way of understanding the role that such degrees could play (if they existed), *and* one that is compatible with the Standard View. As a result, again, we would need a stronger motivation to depart from the Standard View. Simply claiming that there are degrees of awareness isn't enough.²¹

6 Conclusion

My main goal in this paper has not been to give definite answers to any of the questions examined, but to *pose* those questions, to highlight their importance, and the potential importance of bringing together the literatures on permissible defense and responsibility in answering those questions. As we have seen, more work remains to be done to determine whether the concept of responsibility presupposed by Responsibility Views is compatible with the notion that is the focus of the more general responsibility debates. The main problem I've identified is that the neutral

²⁰ On the other hand, if one attempted to revise the Standard View by appealing to differences in kind (not mere differences in degree), that would result in a more radical departure from the original version, which would be harder to motivate. (Of course, I'm not saying that this is impossible, just harder to do. Thanks to Helen Frowe for discussion.)

²¹ Once again, it's important to bear in mind that the relevant concept of responsibility is moral: proponents of the Responsibility View want to claim that the conscientious driver is *morally responsible* for the wrongful threat she poses, despite not being blameworthy for it to any degree. But, on the face of it, it's not at all obvious how this claim could be motivated as a claim about moral responsibility in particular. (In contrast, I can be legally responsible for wrongful acts done by my underage children without being blameworthy for them. But it's not at all obvious how I can be morally responsible for them if I'm not blameworthy to any degree.)

concept presupposed by Responsibility Views is in tension with the Standard View, a well-established and initially plausible general account about the connection between the neutral and non-neutral forms of responsibility. I've drawn attention to a particular strategy that could be used to address this challenge, but I've also pinpointed some important obstacles that stand in the way.

Acknowledgements Thanks to Tyler Doggett, Helen Frowe, Massimo Renzo, the participants at the Conversations on War workshop (Perast, 2019), and two anonymous referees. Special thanks to Helen and Massimo for organizing the workshop and putting together this special issue. This paper was written as part of the Conversations on War workshop series (co-hosted by the Stockholm Centre for the Ethics of War and Peace, Stockholm University and the Yeoh Tiong Lay Centre for Politics, Philosophy and Law, King's College, London).

References

- Alexander, L., & Ferzan, K. (2009). *Crime and culpability*. Cambridge: Cambridge University Press.
- Björnson, G. (2017). Explaining (Away) the epistemic condition on moral responsibility. In P. Robichaud & J. Wieland (Eds.), *Responsibility: The epistemic condition* (pp. 147–162). Oxford: Oxford University Press.
- Capes, J. (2012). Blameworthiness without wrongdoing. *Pacific Philosophical Quarterly*, 93(3), 417–437.
- Coates, J., & Swenson, P. (2013). Reasons-responsiveness and degrees of responsibility. *Philosophical Studies*, 165(2), 629–645.
- Coates, J., & Tognazzini, N. (2012a). The nature and ethics of blame. *Philosophy Compass*, 7(3), 197–207.
- Coates, J. and N. Tognazzini. 2012b. "The Contours of Blame." In J. Coates and N. Tognazzini (eds.), *Blame: Its nature and norms* (Oxford University Press, Oxford), chapter 1.
- Copp, D. (1997). Defending the principle of alternative possibilities: Blameworthiness and moral responsibility. *Nous*, 31, 441–456.
- Doggett, T. (2011). Recent Work on the Ethics of Self-Defense. *Philosophy Compass*, 6(4), 220–233.
- Ferzan, K. (2005). Justifying self-defence. *Law and Philosophy*, 24, 711–749.
- Fischer, J. (2004). Responsibility and manipulation. *The Journal of Ethics*, 8(2), 145–177.
- Fischer, J., & Ravizza, M. (Eds.). (1993). *Perspectives on Moral Responsibility*. Ithaca: Cornell University Press.
- Fischer, J., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. New York: Cambridge University Press.
- Guerrero, A. (2017). Intellectual difficulty and moral responsibility. *Robichaud and Wieland, 2017*, 199–218.
- Haji, I. (1998). *Moral appraisability: Puzzles, proposals and perplexities*. New York: Oxford University Press.
- Hosein, A. (2017). Responsibility and self-defense: can we have it all? *Res Publica*, 23(3), 367–385.
- Kaiserman, A. Forthcoming. Reasons-Sensitivity and Degrees of Free Will. *Philosophy of Phenomenological Research*.
- Kaufman, W. (2009). Can 'Moral Responsibility' Explain Self-Defense? *Robinson, Garvey, and Ferzan, 2009*, 400–402.
- King, M. (2015). Manipulation arguments and the moral standing to blame. *Journal of Ethics and Social Philosophy*, 9(1), 1–20.
- Lazar, S. (2009). Responsibility, risk, and killing in self-defense. *Ethics*, 119, 699–728.
- Lazar, S. (2010). The responsibility dilemma for killing in war: A review essay. *Philosophy and Public Affairs*, 38(2), 180–213.
- McKenna, M. (2012). *Conversation and responsibility*. New York: Oxford University Press.
- McMahan, J. (1994). Self-defense and the problem of the innocent attacker. *Ethics*, 104, 252–290.
- McMahan, J. (2009). *Killing in war*. New York: Oxford University Press.
- Nelkin, D. (2016). Difficulty and Degrees of Moral Praiseworthiness and Blameworthiness. *Nous*, 50(2), 356–378.
- Nozick, R. (1974). *Anarchy, State, and Utopia*. New York: Basic Books.

- Otsuka, M. (1994). Killing the innocent in self-defense. *Philosophy and Public Affairs*, 23, 74–94.
- Otsuka, M. (2016). The moral-responsibility account of liability to defensive killing. In C. Coons & M. Weber (Eds.), *The ethics of self-defense* (pp. 51–68). Oxford: Oxford University Press.
- Pereboom, D. (2001). *Living without free will*. Cambridge: Cambridge University Press.
- Robichaud, P., & Wieland, J. (Eds.). (2017). *Responsibility: The epistemic condition*. Oxford: Oxford University Press.
- Robinson, P., Garvey, S., & Ferzan, K. (Eds.). (2009). *Criminal law conversations*. New York: Oxford University Press.
- Rodin, D. (2002). *War and self-defense*. New York: Oxford University Press.
- Sartorio, C. (2016). *Causation and free will*. Oxford: Oxford University Press.
- Sartorio, C. Forthcoming. Responsibility and causation. In D. Nelkin and D. Pereboom (eds.), *The Oxford handbook of moral responsibility* Oxford: Oxford University Press.
- Smith, H. (1983). Culpable ignorance. *Philosophical Review*, 85, 449–487.
- Strawson, P. (1962). Freedom and resentment. *Proceedings of the British Academy*, 48, 1–25.
- Tadros, V. (2018). Causal contributions and liability. *Ethics*, 128(2), 402–431.
- Thomson, J. (1991). Self-defense. *Philosophy and Public Affairs*, 20, 283–310.
- Tierney, H. (2019). Quality of reasons and degrees of moral responsibility. *Australasian Journal of Philosophy*, 97(4), 661–672.
- Usher, M. (2020). Agency, teleological control, and robust causation. *Philosophy and Phenomenological Research*, 100(2), 302–324.
- Wallace, J. (1994). *Responsibility and the moral sentiments*. Cambridge, MA: Harvard University Press.
- Watson, G. (1996). Two faces of responsibility. *Philosophical Topics*, 24(2), 227–248.
- Widerker, D. (1991). Frankfurt on 'ought implies can' and alternative possibilities. *Analysis*, 51, 222–224.
- Zimmerman, M. (1988). *An essay on moral responsibility*. Totowa: Rowman and Littlefield.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.