



Manipulation and liability to defensive harm

Massimo Renzo¹

Accepted: 18 January 2021
© The Author(s) 2021

Abstract Philosophers working on the morality of harm have paid surprisingly little attention to the problem of manipulation. The aim of this paper is to remedy this lacuna by exploring how liability to defensive harm is affected by the fact that someone posing an unjust threat has been manipulated into doing so. In addressing this problem, the challenge is to answer the following question: *Why should it be the case (if it is, indeed, the case) that being misled into posing an unjust threat by manipulation makes a difference to one’s liability, as compared to being misled into doing so by natural events or by someone’s honest attempt to persuade us?* To answer this question, I first outline an account of manipulation and then use it to defend what I shall call the “Pre-emption Principle.”

Keywords Liability · Defensive harm · War · Manipulation · Reasons · Practical deliberation

1 Manipulation in war

Manipulation is a pervasive phenomenon in war, raising important questions both at the *ad bellum* and *in bello* stages. Consider the *ad bellum* stage first. Wars can be successfully waged only if there is enough support from the armies that are supposed to fight them and the populations that will have to bear the various costs associated with them. For this reason, the real aims of war are sometimes intentionally concealed by governments, particularly when they are such that support for the war could not be easily secured if they were revealed. For example, wars that in fact aim to secure financial or strategic advantages are presented as justified by the need to neutralize

✉ Massimo Renzo
massimo.renzo@kcl.ac.uk

¹ King’s College London, London, UK

existential threats or protect values such as freedom, democracy or human rights. Moreover, whatever their declared goals, governments have historically resorted to various forms of propaganda to drum up support for their wars by playing with the beliefs, emotions and attachments of their citizens.

At the *in bello* stage, a number of manipulative behaviours are in fact permitted (at least legally), including feigned attacks or retreats, transmitting misleading radio messages, spying, intentional planting of false information, and using dummy installations to intimidate as well as to encourage useless attacks. The permission to resort to manipulation, however, is limited to behaviours that fall short of perfidy. The laws of war rule out, among other things, pretending to surrender in order to secure an advantage over the enemy and abusing the protections afforded to medical personnel (say, by disguising combatants as doctors or members of the Red Cross).¹ What counts as perfidy is not always entirely clear, but it is accepted that there is a line and that it is important to respect it (Ripstein 2020).

Yet, philosophers working on war and, more generally, on the morality of harm have paid surprisingly little attention to the problem of manipulation.² It is routinely mentioned that being the victim of manipulation might be an excuse for posing an unjust threat, thereby undermining one's responsibility, but there is no sustained discussion in the literature of why exactly manipulation excuses and how, nor is there a discussion of the distinction between excuses associated with coercion, on the one hand, and manipulation, on the other. The aim of this paper is to remedy this lacuna by exploring the question of how liability to defensive harm is affected by the fact that the person posing (or contributing to) an unjust threat has been manipulated into doing so.

One problem in addressing this question is that whilst philosophers have written extensively about coercion, manipulation has received scant attention in the philosophical literature,³ so our understanding of what manipulation is, or why it is wrong to manipulate people (when it is wrong to do so), is somewhat limited. To further complicate things, manipulation does not seem to be a single moral phenomenon. Rather, manipulation can take different forms, as the examples above illustrate.

Consider three cases:

- The use of human shields;
- The brainwashing and recruitment of child soldiers;
- The use of deceptive strategies, such as those in which false information is intentionally planted to encourage the enemy to behave in certain ways.

These cases are structurally very different. The first involves manipulating the circumstances so as to generate new reasons for the manipulee to act as the

¹ Article 23 of the 1907 Hague Convention IV. See also Protocol I of the Geneva Conventions, Articles 37, 38, and 39.

² A recent exception is Fabre (2020), where the problem is addressed in the context of a discussion of the morality of spying.

³ Recent works include Noggle (1996), Baron (2003), Greenspan (2003), Buss (2005), Gorin (2014a) and the articles collected in Coons and Weber (2014).

manipulator wishes. Placing civilians in front of soldiers who are shooting at their enemy, for example, creates new reasons for the enemy not to shoot back, since doing so would involve harming civilians (Fabre 2012, 256–267; Haque 2017, chap. 9).

The distinctive feature of the second case, by contrast, is that instead of acting on the circumstances, the manipulator acts on the values and the goals of the manipulee. When a child soldier is brainwashed into thinking that the members of a certain minority lack the sort of moral value that generate deontological restrictions against harm, she is induced to act as the manipulator wishes because certain values or goals have been implanted in her in a way that somehow bypasses her capacity to decide for herself how to act.⁴

Finally, cases of the third kind are those in which the manipulator acts neither on the circumstances in which the manipulee find herself, nor on the values and goals that should guide her choice, but rather on the manipulee's perception of what she has reason to do in light of those values, goals and circumstances. This can be achieved in two ways: The manipulator might induce the manipulee to make a mistake of fact—for example, induce her to mistakenly believe that a certain area is not densely populated, or that a given building is a military factory, rather than a hospital. Alternatively, the manipulator might induce the manipulee to make a moral mistake—for example, induce her to mistakenly believe that punitive torture is not incompatible with her own moral goals and the values that should guide her conduct.⁵

I doubt that these are all *species* of the same *genus*. There seem to be important differences between them, and thus we should not assume that they are all subject to the same kind of analysis. For the purposes of this paper, I will focus exclusively on cases of the third type.⁶ So consider, for example:

Bomb: Al provides overwhelming evidence to Bianca of the fact that Clint is about to detonate a bomb that will kill a large number of innocent people. Al knows that Clint is not about to detonate any bomb, but he intentionally misleads Bianca because of his dislike of Clint. Bianca can kill Clint at no

⁴ These cases have been extensively debated in the free-will literature, in relation to the question of whether responsibility is compatible with determinism, as well as with the sort of indeterminism implied by standard versions of quantum mechanics (Pereboom 2014; Fischer 2017; Mele 1995). I will not consider this question here.

⁵ The distinction between cases in which one's values and goals are manipulated (as in the brainwashing of child soldiers) and cases in which what is manipulated is one's perception of what her own values and goals require can be elusive, and no doubt there will be cases whose classification is unclear. Still, the distinction is significant. In one case, the manipulator implants in his victim new goals and values; in the other, the manipulator takes the goals and values of his victim for what they are and confuses her instead about what they require of her.

⁶ Cases of the first kind, sometimes called "moral coercion" cases, are discussed in Bazargan (2014b) and Frowe (2020). It is only on a broad understanding of "manipulation" that they count as manipulative (Wood 2014), and this classification is not uncontroversial (Baron 2003). Indeed, manipulation is sometimes defined by contrasting it with coercion on the one hand, and rational persuasion on the other (Blumenthal-Barby 2012).

significant cost to herself, and thus reasonably believes that she has a duty to do so.

The questions we shall consider are: Is Bianca liable to be killed in defense of Clint? Under what conditions? And what difference, if any, does it make to her liability that she was manipulated by Al? Would Bianca's liability be different if her mistake had not been the product of manipulation?

2 The challenge

In providing an account of how manipulation affects liability to defensive harm, the challenge is this: Why should it be the case (if it is, indeed, the case) that being misled into ϕ -ing by manipulation makes a difference to one's liability, as compared to being misled into ϕ -ing by natural events or by someone's honest attempt to persuade us that ϕ -ing is what we have reason to do?⁷

To see this, compare *Bomb* with

Bomb II: Bianca has overwhelming evidence that Clint is about to detonate a bomb that will kill a large number of innocent people. Bianca's belief is not unreasonable, and she has not been negligent in collecting evidence. However, Bianca's belief is mistaken. In fact, Clint is not about to detonate a bomb. Bianca's evidence is the product of natural causes. (Say, an optical illusion produced by the odd lighting in the building.) Bianca can kill Clint at no significant cost to herself, and thus reasonably believes that she has a duty to do so.

and

Bomb III: Like *Bomb*, except here Al genuinely believes that Clint is about to detonate the bomb. Acting in good faith, Al persuades Bianca that this is the case. Al's belief is not unreasonable, and he has not been negligent in collecting evidence. However, Al is mistaken, and Clint is innocent.

If liability is affected by manipulation, we should expect Bianca's liability in *Bomb* to be different from her liability in *Bomb II* and *Bomb III*. Is that really the case? And if so, why? After all, in all three cases, Bianca autonomously decides how to act based on the evidence available to her. Holding everything else equal, why should it matter how the evidence she is responding to has been produced? No matter what the causal history of that evidence, we might think, what determines Bianca's liability is how she decides to act on the basis of it.

This is particularly clear if we focus on *Bomb* and *Bomb III*. Since in *Bomb*, Bianca is not aware that Al is trying to manipulate her, from her perspective *Bomb*

⁷ Since I am interested in the question of how liability for contributing to an unjust threat is affected by manipulation, I will focus on cases like *Bomb*, in which ϕ -ing is not what B has reason to do. Of course, one might be manipulated into ϕ -ing where she does have reason to ϕ . (For example, the kidnapper might be manipulated by the police into releasing the hostages).

and *Bomb III* are indistinguishable. Why then think that that her liability might differ between these cases?

Of course, our answer to this question will depend on what we think the basis of liability to defensive harm is. Here I will focus on the idea that liability is grounded in moral responsibility for posing an unjust threat. I do so not because I think that this idea provides an exhaustive answer to the question of the justification of liability—plausibly, liability is also determined by other considerations—but rather because it is widely considered a central component of the justification of liability, even by those who deny that it is the only component (McMahan 2009; Otsuka 1994; Tadros 2020b). Thus, although a full account of how manipulation affects liability will have to consider how other potential grounds for liability (such as culpability, complicity, or attempting but failing to pose a threat) are affected by manipulation, given the prominence of the Responsibility View, this is a natural place to start in investigating the issue.⁸

Now, at least some of those who accept that liability is grounded in responsibility will be tempted to conclude that Bianca is not liable in *Bomb*, since her responsibility is undermined by the fact that she is the victim of manipulation. We are *morally*, as opposed to *causally*, responsible for ϕ -ing only if ϕ -ing is the product of how we have exercised our autonomous agency. But manipulation is normally thought to prevent us from acting as autonomous agents, thereby undermining our responsibility. When we ϕ as a result of manipulation, there is an important sense in which our ϕ -ing cannot be properly attributed to us. And if it cannot be attributed to us, it seems plausible that we are not morally responsible for ϕ -ing. So, we might think that if moral responsibility is what grounds liability, whenever responsibility is undermined, liability is undermined as well.

On the face of it, this strategy has some plausibility. Many, for example, will describe Bianca's situation in *Bomb* as one in which she is treated like a "tool" or an "instrument" in Al's hands. Bianca does, of course, retain the capacity to reason and deliberate about what to do. But insofar as Al pulls her "psychological levers" (Noggle 1996), there seems to me a sense in which she is prevented from making up her own mind about what to do, and thus she is prevented from acting "for her own reasons."

Of course, Bianca does engage in a process of deliberation when she decides to attack Clint, but because her decision is based on Al's misrepresentation of what she has reason to do, Bianca is not treated by Al as a "possible source of reasons all the way down" (Herman 1993). Indeed, some may go as far as to say that, effectively, Al has "taken over her agency", so that the killing of Clint in *Bomb* is in fact best regarded as Al's act. After all, the outcome of Bianca's process of deliberation is ultimately determined by Al in precisely the way he had intended. By contrast, Bianca fails to even understand that she is attacking an innocent person. If so, how can this choice be attributed to her in a normatively significant sense? Following Korsgaard, we might be tempted to say that Bianca "cannot 'contain in [her]self"

⁸ The view defended in this paper is also compatible with other accounts of liability, for example the Duty View (Tadros 2013).

the end of the same action because [she] is not in a position to. [She] doesn't know what the real end of the action is, and is therefore not in a position to make it [her] own—to choose, freely, to contribute to its realization” (Korsgaard 1996).

The problem with this strategy is that it is not clear in what sense Bianca's reasons are not “her own”, or in what sense she is not the source of those reasons “all the way down”. In a way, we are never the source of our reasons for action “all the way down”. We respond to features of the world as it presents itself to us. Our choice to ϕ is never made out of thin air. The reasons that we have to ϕ depend on the circumstances we find ourselves in and on our reading of those circumstances.

There is, of course, another sense in which it might be said that we are not “acting for our own reasons” when we decide to ϕ as a consequence of manipulation: Φ -ing is not what we would do if we had a better understanding of our circumstances. But this is always the case when we deliberate about how to act. Even when we are not manipulated, it is often true that we would have acted differently, had the evidence available to us been more accurate. (We can assume, for example, that Bianca would have refrained from attacking Clint in *Bomb*, had she had better evidence of his innocence. But plausibly the same assumption would hold in *Bomb II* and *Bomb III*.)

Sarah Buss (2005) powerfully makes this point. It is misleading, Buss argues, to think that in cases of manipulation, our autonomous agency is undermined in virtue of the fact that our conduct cannot be attributed to us. It can. After all, we would have acted in the same way if our response to the evidence in question had been prompted by natural events. (See also, Yaffe 2003).⁹

This naturally leads to the rejection of the view that manipulation undermines liability. If manipulation does not prevent us from acting as autonomous agents, then our ϕ -ing can be attributed to us. But if ϕ -ing can be attributed to us, then we're responsible for ϕ -ing. And if responsibility is what grounds liability, it looks as if we should conclude that we are liable to bear the costs associated with ϕ -ing, even when we have been manipulated to ϕ .

This conclusion, however, is hard to accept. The intuition that Bianca's liability in *Bomb* is affected by the fact that she has been manipulated by AI does seem to retain some force. Whether she is liable to be defensively harmed, and to what extent, seems to depend, at least to some extent, on whether her conduct is the product of AI's manipulation or not. Do we need to abandon this intuition if we are to accept Buss' persuasive argument that Bianca is no less capable of exercising her autonomous agency in *Bomb* than she is in *Bomb II*?

I argue that we need not. I will defend the view that liability is indeed affected by manipulation, while conceding both that our responsibility is grounded in our capacity to exercise our autonomous agency and that manipulation does not undermine such capacity (or at least that it doesn't do so any more than other mistakes we make without being intentionally misled by others). Moreover, in

⁹ This is a well-known problem in the free will literature. How can we distinguish between the influence of manipulation, which we tend to regard as autonomy-undermining, and the influence of natural events, which we are not inclined to consider autonomy-undermining? (Pereboom 2014; Fischer 2017; Mele 1995).

developing this argument I will not abandon the idea that it is responsibility for posing unjust threats that (among other things) grounds liability to defensive harm.

In other words, I will concede that Bianca's responsibility in *Bomb* is no different from her responsibility in *Bomb II* and *Bomb III*, and yet argue that her liability is.¹⁰ This is because in cases like *Bomb*, the liability of the manipulee is pre-empted by the liability of the manipulator. The degree to which liability is pre-empted will vary depending on a number of variables, but there will be cases in which only the manipulator is liable to defensive harm and the manipulee is not. To make my case, I will introduce and defend what I shall call the "Pre-emption Principle."

3 The Pre-emption Principle

Manipulating someone involves interfering with their capacity to respond to reasons. Qua rational agents, we have the capacity to recognize certain features of the world as salient reasons for action. When we are manipulated, the psychological processes through which we respond to these reasons are compromised in one of two ways: Either we are intentionally led to fail to do what we have reason to do (for example, you might exploit my insecurity to induce me to oppose anti-discrimination policies in the office) or we are intentionally led to do what we have reason to do for the wrong reasons (you might persuade me to support anti-discriminatory policies, but to do so only because my boss will be favourably impressed).

Importantly, our reasons for action are typically mediated by the reasons we have to form certain beliefs and attitudes. Normally, we form the intention to ϕ because we believe that ϕ -ing is what we have reason to do, or because ϕ -ing is called for in light of certain attitudes—for example, certain desires or emotions—of ours. Manipulation involves interfering with our capacity to respond to our reasons for action by compromising our capacity to respond to our reasons for beliefs and attitudes in a distinctive way.¹¹

¹⁰ One way to resist the claim that Bianca can retain responsibility for her conduct despite having been manipulated is to point out that she is not aware of what she is doing. Since manipulated agents fail to satisfy the so called "epistemic condition for moral responsibility," the argument goes, they cannot be considered responsible for what they do. (Thanks to Carolina Sartorio for pressing this point.) This line of argument, however, is not going to help here. For Bianca is no more aware of what she does in *Bomb II* and *Bomb III* than she is in *Bomb*. If her ignorance is what undermines responsibility, thereby undermining her liability, we would have to conclude that she is not responsible, and therefore not liable, in all of these cases. The alleged asymmetry between the first two cases and the third one would not be vindicated.

¹¹ For the purposes of this paper, I will focus on how manipulation undermines our capacity to respond to reasons for action by interfering with our capacity to respond to our reasons for beliefs and attitudes. But there can be instances of manipulation that involve interfering with our capacity to respond to our reasons for beliefs and attitudes *per se*. I might manipulate you into forming a mistaken belief or to feel emotions that are uncalled for, without this being part of an attempt to compromise your capacity to respond to reasons for action. Different articulations of the idea that manipulation involves undermining our capacity to respond to reasons can be found in Mills (1995), Gorin (2014b), and (Hannah 2015). An influential alternative account that appeals to the idea that manipulation involves attempting to lead someone to violate certain ideals is offered in Noggle (1996).

After all, I could get you to fail to act upon the reasons for action that apply to you simply by depriving you altogether of the capacity to exercise your deliberative faculties.¹² I could intoxicate you or hit you in the head. These obviously would not constitute instances of manipulation. Manipulation involves exploiting your own deliberative faculties in order to produce the relevant disconnect between the reasons that ought to move you and the reasons that do move you. *You* are the one who comes to decide that you ought to oppose antidiscrimination policies, though you're led to form this decision by the misleading clues and arguments I have produced (and that I have produced precisely with the intention of leading you to make that decision). Just like Tom Sawyer's friends decide themselves that they should spend the afternoon painting Aunt Polly's fence, though they are expertly drawn to make this decision by Tom's feigned enthusiasm (which he has feigned precisely with the intention of leading them to make the decision).

The reasons why the wrong of manipulation is so insidious is precisely that it is by exploiting our own deliberative faculties that the manipulator undermines our capacity to exercise our own practical agency. In exercising our practical agency, our aim is to do what we have reason to do and, ideally, to do it for the right reasons. The manipulator intentionally misleads us about what those reasons are, so that the more we try to respond to what we have reason to do, the more we fail to do so.¹³

One way to put this idea is to say that the manipulator intentionally causes his victim to have apparent reasons for action (Parfit 2011, 111) that are misaligned with her actual reasons, and that he achieves this result by intentionally causing his victim to have apparent reasons to form belief or attitudes that are misaligned with her actual ones. For example, in *Bomb*, Al intentionally causes Bianca's apparent reasons for action to be misaligned with her actual ones by producing apparent reasons for belief that do not track her real reasons. Bianca ends up failing to act upon her reasons for action because Al has intentionally produced apparent reasons that are at odds with what (Al knows) she has reason to do.¹⁴

Different kinds of manipulation will employ different means (deception, psychological pressure etc.) to produce this result. Since we're interested in cases like *Bomb*, let us focus on deceptive manipulation.

Deceptive Manipulation: A deceptively manipulates B when A uses deception to intentionally interfere with B's deliberative faculties (in the way described above), so that B will fail to act upon her reasons for action in one of two ways:

- a. B has reason to not ϕ and A's interference intentionally causes B to ϕ ;
- b. B has reason to ϕ and A's interference intentionally causes B to ϕ for the wrong reasons.

¹² "Deliberative faculties" here is to be understood in a broad sense, as referring to the psychological processes through which we respond to reasons in general, including reasons for beliefs and attitudes.

¹³ This is the outline of an account of manipulation that I develop elsewhere.

¹⁴ Henceforth, I will reserve the term "reasons" without qualifications for our actual reasons, as opposed to apparent ones. I will also assume that, despite her epistemic limitation, Bianca retains reasons not to harm Clint.

Two clarifications are in order: First, I will assume that the misalignment between the victim's actual reasons for action and her apparent ones is the product of manipulation only if the interference in question is intentional. When it isn't, although we might be misled by someone else's conduct, we are not the victims of manipulation. Suppose, for example, that the maid accidentally misplaces Desdemona's handkerchief, putting it in Cassio's room, thereby leading Othello to believe that Desdemona is unfaithful. We would not say that the maid has thereby manipulated Othello.¹⁵

Second, for our purposes, we can focus here on cases of manipulation that involve intentionally causing someone to do what they have reason not to do (clause *a*), leaving aside cases in which one is intentionally led to do what she has reason to do for the wrong reasons (clause *b*). This is because liability to defensive harm is typically at issue when someone is doing what she has reason not to do, i.e. posing a wrongful threat. (In *Bomb*, for example, Bianca has reason to refrain from killing Clint. In misleading her to kill Clint, Al intentionally leads her to do what she has reason not to do.)

With this, I do not mean to rule out the possibility of cases in which liability is triggered by the fact that the agent acts as she should but for the wrong reasons. Perhaps cases like this exist. Consider:

Tea Break: Larry intends to kill Teo by poisoning his tea, but instead of pouring poison, he accidentally pours some vitamins in the cup. Teo is aware of that. At the same time, Paulie, who is not acting in coordination with Larry, shoots Teo. Teo can only defend himself by using Larry as a shield.

I think it's plausible that Larry is liable to be killed by Teo, even if Larry is conforming with his reasons not to harm Teo (Tadros 2013), in virtue of the fact that he does so for the wrong reasons.¹⁶ At any rate, these cases are significantly less common than cases like *Bomb*, and thus I will not discuss them here.

So, how does manipulation, understood in the way suggested above, affect Bianca's liability? I will defend the following principle:

When B ϕ -s and the following conditions are in place:

- i. B ϕ -s in response to apparent reasons that are misaligned with her actual reasons;
- ii. the misalignment between the actual reasons and the apparent ones is intentionally produced by A's interference with B's practical deliberation in the way described above;
- iii. A acts culpably or recklessly;

¹⁵ But see Manne (2014) for a criticism of this view.

¹⁶ On the distinction between "conforming" and "acting upon" (or "complying with") our reasons for action, where the latter involves not only acting as we have reason to, but also being motivated to do so by the right reasons, see Raz (1999).

A, rather than B, is liable to bear (at least some of) the costs associated with B ϕ -ing, provided that doing so will secure whichever aims justify the imposition of those costs.¹⁷

Call this the “Pre-emption Principle,” as it states that under conditions (i) to (iii), A’s liability pre-empts B’s liability.

Why is A liable to bear some of the costs associated with B’s ϕ -ing when these conditions are met? When B is manipulated by A into committing an act of wrongdoing, B is severely wronged by A. Not only is B prevented from acting upon the reasons that apply to her,¹⁸ but because of this failure on her part, B violates her duty not to harm innocents.

B is wronged by A simply in virtue of the fact that she is prevented from responding to her reasons for action as she should. But the wrong is compounded by the fact that in failing to do so, B violates a serious duty. This raises the question of who should bear the costs associated with preventing the violation of this duty, or at least mitigating its consequences. The answer is that since A’s intervention is what caused B’s failure, and A is morally responsible for so intervening, A incurs a secondary duty to bear those costs (Raz 2011; Gardner 2018; Tadros 2020a).

So, for example, in *Bomb*, Al had a primary duty not to intentionally get Bianca’s apparent reasons to be misaligned with her actual reasons. Having breached that primary duty, Al incurs a secondary duty to make up for it. How can Al make up for it? He can do so by bearing the relevant costs associated with the fact that Bianca has been intentionally misled. The relevant costs, in this case, are those associated with Clint’s defensive attack.

A natural way to think about the Pre-emptive Principle is in terms of a duty to protect Bianca that Al incurs by manipulating her. The remedial duty incurred by Al is a duty to protect Bianca from Clint’s defensive attack; and to do so at the cost of his life, if the alternative is Bianca’s death. This is because protecting Bianca is what it takes to offset the costs imposed on her by manipulating her. And since Al is under a duty to protect Bianca at the cost of his life, others are also required, all things equal, to direct any defensive attacks (including lethal ones) against Al.

So, suppose that Clint could defend himself by killing either Al or Bianca. The Pre-emption Principle tells us that Clint ought to choose Al. Bianca would be wronged if Clint were to kill her, rather than Al. He might, of course, be excused for doing so, if he’s not aware that Bianca has been manipulated by Al. But if Clint tries to defend himself by killing Bianca (say, because he’s not aware of Al’s plot), Bianca is permitted to defend herself by diverting Clint’s defensive harm toward Al,

¹⁷ I am assuming that B does not bear responsibility for his condition, in the sense that B did not culpably, recklessly or negligently expose himself to the risk of manipulation. B might retain some liability if tracing considerations of this kind apply (Rosen 2003; Fischer and Tognazzini 2009).

¹⁸ We have seen that when B is manipulated by A, B’s capacity to act upon the reasons for action that apply to her is precluded in the distinctive way discussed at the beginning of this section. For ease of exposition, I shall take this qualification for granted in the rest of the paper.

or even by using AI as a shield. AI, not Bianca, is liable to be killed defensively by Clint.¹⁹

It is true, of course, that Bianca might have failed to respond to her reasons for action, and thus might have violated her duty not to harm Clint, even without AI's intervention. Natural causes could have produced the misleading evidence (as in *Bomb II*). But we need to tread carefully in drawing conclusions from this observation.

Absent further details, the answer to the question of whether Bianca is liable in *Bomb II* is not straightforward because, as we have seen, responsibility is not the only criterion of liability. Suppose, however, that we fill in the details of the case so that Bianca's responsibility is sufficient to make her liable to be killed defensively to save Clint. (Those who adopt the Responsibility View might say, for example, that Bianca is liable insofar as the unjust threat of harm to Clint has arisen as a result of her choices, and it was foreseeable for Bianca—though not likely, given the evidence available to her—that she might end up posing an unjust threat; McMahan 2009; Otsuka 1994). This does not change the fact that in a scenario in which AI is responsible for manipulating Bianca she is not liable in the same way. Since in this scenario, AI has a duty to bear at least some of the costs of Clint's defensive attack, Bianca would be wronged if she were made to bear those costs instead. And to say

¹⁹ Does this mean that Clint is liable to be killed by Bianca? Not necessarily. Epistemic considerations seem to play a role here. For example, it makes a difference whether Clint is aware that AI is liable and Bianca isn't. Suppose that Clint could defend himself by killing either AI or Bianca. If he is aware that AI is liable and Bianca isn't, but he targets Bianca, it seems plausible that he is liable to be killed defensively by her. What if the evidence available is such that Clint could not be aware of that? It's unclear how we should deal with cases like this. From the fact-relative perspective, killing Bianca is unnecessary. (On the distinction between "fact-relative" and "evidence-relative" perspective, see Parfit 2011.) Given that Clint could survive by killing the liable party, killing the non-liable one does not constitute the least harmful means to defend himself. If so, we might think that Clint becomes liable to defensive harm by wrongfully attacking Bianca. Alternatively, we might think that since Clint could not possibly know that AI is liable and Bianca isn't, targeting Bianca does not violate the necessity constraint, and thus Clint is not liable.

We need not take a position on this problem here, since it does not affect the truth of the Pre-emption Principle. For the role of the Pre-emption Principle is to explain how our liability to defensive harm is affected when we are either responsible for, or victim of, manipulation, and Clint is neither. People holding different views about how his ignorance should affect his liability can thus accept what the Pre-emption Principle tells us about AI's and Bianca's liability. The same problem, after all, arises in cases that do not involve manipulation. Consider:

Trolley: Bystander can divert a trolley away from five towards either Workman or Engineer. Unbeknown to Bystander, Engineer culpably set the trolley in motion. Bystander tosses a coin and, as a result of the toss, moves to divert the trolley towards Workman.

If Bystander knew all the facts, it would be wrong of her to divert the trolley towards Workman. Engineer is liable to have the trolley diverted toward him, whereas Workman isn't. But what if Bystander doesn't know all the facts, and her evidence is that she ought to toss a coin? Suppose that Workman could defend herself by harming Bystander defensively (without thereby preventing the saving of the five). Is Bystander liable to such harm? Here too, how we answer this question depends on whether we understand the necessity constraint in fact-relative or evidence-relative terms. (See Lazar (2012) for related discussion, though Lazar focuses on cases in which what the victim of the attack is unaware of are non-normative facts—for example, that there is an escape route that would enable her to survive without harming the attacker. As far as I know, cases in which the victim is unaware of normative facts, such as the attacker's liability, have not been discussed in the literature.) Thanks to Helen Frowe for helping me see this issue more clearly.

that Bianca would be wronged if she were made to bear those cost, is just to say that she is not liable to bear them.

It is often noticed that whether harming someone to avert a threat wrongs her depends on whether it would be fair to impose those costs on her vis-à-vis other parties (McMahan 2005; Draper 2016; Tadros 2020b). But generally, those who make this point have in mind cases in which the unfairness in question depends on the fact that the party made to bear the costs is overall less responsible (or perhaps less culpable) than the others. There are, however, further reasons why it might be unfair to make B, rather than A, bear the costs of averting a threat they are both posing. One such reason is that B has been wrongfully manipulated by A to pose the threat in question, and thus is owed a remedial duty by A. In other words, it's not only the overall level of responsibility of the parties involved that can make it unfair to choose one over the other, but also special claims that the parties might have vis-à-vis each other.²⁰ Being the victim of manipulation can generate such claims.

If so, we can concede Buss' point that Bianca is no less capable of exercising her autonomous agency in *Bomb* than she is in *Bomb II*. We can also concede that in virtue of that, her responsibility in the two cases is the same.²¹ This is still not enough to conclude that Bianca's liability in the two cases is also the same. Since in *Bomb* Al has a duty to bear some of the costs associated with defending Clint, Bianca is not liable to bear those costs in the same way in which she is in *Bomb II*.

Bianca, however, is not liable to bear those costs only provided that Clint can be protected by imposing them on Al. When this is not the case, Bianca *is* liable. This is because what grounds her non-liability is the fact that Al has a remedial duty toward her—a duty which requires Al to protect Bianca by bearing the costs of Clint's defensive attack. It's for this reason that Bianca's liability is pre-empted by Al's. When Clint cannot be defended by killing Al, however, Al cannot discharge his remedial duty by protecting Bianca. Because of this, Bianca's liability is not pre-empted, and thus she ends up fulfilling the conditions of liability in the same way she does in *Bomb II* and *Bomb III*.²²

²⁰ These two considerations don't always go hand in hand. As we will see, there might be cases in which B's responsibility is intentionally produced by A's manipulative interference and yet B is more responsible.

²¹ We'll have to concede that if we believe: (a) that responsibility is grounded in our capacity to exercise our autonomous agency, and (b) that manipulation does not undermine such capacity any more than other mistakes we make without being manipulated. Some might deny this and insist, *contra* Buss, that Bianca's responsibility is undermined in virtue of the fact that she is manipulated. This opens up the possibility of an alternative formulation of the Pre-emption Principle, according to which it is Bianca's responsibility, not simply her liability, that is pre-empted.

²² Why take the view that Bianca's liability is pre-empted in *Bomb*, rather than saying that both Al and Bianca are liable? Following Tadros, we might argue instead that although both Al and Bianca are "individually liable", Al is more "selectively liable" than Bianca, where this means that she would be wronged if Clint were defended by harming her, rather than Al (Tadros 2020b). However, it's unclear what is at stake in labelling Bianca as "individually liable", if she is not selectively liable. Since the point of invoking the notion of liability is to mark the fact that one would not be wronged by the imposition of a given harm, what do we gain by saying that Bianca is (individually) liable, if we also want to say that she *would* be wronged by the imposition of the defensive costs of saving Clint? Better to acknowledge that Bianca's liability is conditional in the way suggested by the Pre-emption Principle.

How shall we determine which costs AI is required to bear according to the Pre-emption Principle? The idea behind the Principle is that we should assign to A liability for the relevant costs associated with his manipulation of B. I suggest that in order to do that, we need first to consider the costs that Bianca would be liable to bear for her conduct if the apparent reasons for action produced by AI's interference had correctly tracked her actual reasons for action. We then treat Bianca as liable to bear those costs, and AI as liable to bear the remaining ones.²³ (For example, in *Bomb*, Bianca would be liable to bear no costs for attacking Clint, if the apparent reasons for action produced by AI's manipulative interference had correctly tracked her actual reasons for action. Thus, Bianca is not liable to bear any cost to defend Clint. AI is liable to bear all of them.)

This is not to say that whenever B's decision to attack V is the product of manipulation by A, B is never liable to defensive harm. B is still liable to bear any costs of her decision to attack V that are generated by how B responds to her apparent reasons. For what grounds the pre-emption of B's liability is A's duty to offset the costs that B would have to bear in virtue of the fact that, because of A's manipulative interference, B's apparent reasons for action do not align with her actual reasons for action. But B might still fail to respond to her apparent reasons. If she does, any liability generated by such failure is not pre-empted. B is liable to bear the costs associated with that failure (absent further excusing considerations).

The idea behind the Pre-emption Principle is that A is liable to bear *the relevant costs*, not *any costs*, associated with A's manipulation. The relevant costs are those that would be unfair to impose on B in light of how she has chosen to respond to her perceived reasons for action. The costs that it is not unfair to impose on B are those associated with how she has chosen to respond to those reasons.

The Pre-emption Principle is thus incompatible with the radical view that manipulation completely eliminates our moral responsibility by precluding the possibility of acting as autonomous agents. But that view seems clearly mistaken. The mere fact that Othello has been manipulated by Iago does not mean that he is not liable to defensive harm to save Desdemona. For even if the facts had been as Othello had reason to believe them to be, given the evidence available to him, attempting to kill Desdemona on those grounds clearly would have made him liable to defensive killing.

So, consider:

Bomb IV: Like *Bomb*, except that Bianca knows that she can kill Clint either by shooting him or by throwing a grenade that, in addition to killing Clint, will also severely harm Dan, an innocent bystander, as a side-effect.

If Bianca chooses to throw the grenade, she is not liable to bear the costs associated with defending Clint. AI is, because of the Pre-emption Principle. However, Bianca *is* liable to bear the costs associated with defending Dan. The Pre-emption Principle does not cover those. So, if Dan could push the grenade away from himself,

²³ Absent any further relevant moral consideration. Harming Bianca might be permissible, for example, on lesser evil grounds.

knowing that this would impose some harm on Bianca (while still saving Clint with the only further cost of Al's life), he would be permitted to do so.

Indeed, there might be even cases where Bianca ends up being liable to more defensive harm than Al. Consider:

Coffee break: Al convinces Bianca that pushing Clint (whom Al dislikes) is necessary to prevent Clint from spilling his coffee on Bianca's new dress, despite knowing that Clint has no coffee in his cup. Instead of pushing Clint, Bianca throws a heavy object at him that will severely injure him.

In this case, Bianca is liable to bear more of the costs associated with defending Clint than Al. For Bianca's level of liability generated by her attempt to severely harm Clint would be significant even if Bianca's apparent reasons for action had correctly tracked her actual ones (i.e. even if Bianca had been right in believing that she had reason to prevent Clint from spilling coffee on her dress).²⁴

4 Refining the Pre-emption Principle

We have seen that the Pre-emption Principle can be understood in terms of a duty to protect Bianca that Al incurs by manipulating her. In this sense, the Principle fits nicely with the Duty View (Tadros 2013), according to which those responsible for creating a wrongful threat to innocent parties can be made to bear the defensive costs required to prevent the threat in question from being realized. But notice that the Pre-emption Principle is not, strictly speaking, a principle for the distribution of defensive harm. It is a broader principle for the allocation of remedial duties generated by manipulative behaviour.

In *Bomb*, the Principle requires Al to protect Bianca from defensive harm. But there will be cases in which the costs in question will be of a different kind. For example, if A has manipulated B into purchasing a needlessly expensive insurance policy, A has a remedial duty to compensate B for the extra costs incurred. If A has manipulated B into becoming estranged from her family, A has remedial duty to facilitate a reconciliation between them. In all these cases, A has wronged B by interfering with her practical deliberation in the way described above. The best way to remedy that wrong is for A to offset the costs that B has to bear because of how her deliberation has been impaired. A can do that by taking on those costs himself.

This difference between the Pre-emption Principle and the Duty View is important even when we restrict our attention to cases of defensive harm. The focus

²⁴ Indeed, depending on the details of the case, some might doubt that Al bears any liability to defensive harm to protect Dan. Whether he does will depend, at least in part, on whether Al could be reasonably expected to foresee that Bianca might overreact and pose a significant threat to Dan when misled by Al's manipulation. If he isn't, the mere fact that her threat to Dan is causally downstream from Al's wrongful manipulation might be deemed insufficient to make Al liable to defensive harm, though it might be enough to ground a permission for Dan (or someone acting on his behalf) to kill Al, if necessary. I shall not address these complications here, as doing so would require adopting a particular view of how exactly responsibility grounds liability, especially in relation to so-called "minimally responsible threateners" (McMahan 2009; Bazargan 2014a).

of the Pre-emption Principle is not on the remedial duties that wrongdoers owe the intended victims of their attack, but rather on the duties they owe the victims of manipulation. And although there can be cases in which the two will coincide (most clearly, when A manipulates B into self-harming), they need not. Indeed, accepting the Pre-emption Principle does not commit one to accepting the Duty View. My strategy in this paper, for example, has been to presuppose the Responsibility View, rather than the Duty View, as the correct account of liability to defensive harm, and then appeal to the Pre-emption Principle to explain how our liability is affected when we are responsible for, or victims of, manipulation.

But doesn't the Responsibility View already have the resources to deal with cases of manipulation without invoking the Pre-emption Principle? Couldn't we conclude that Al, rather Bianca, is liable in *Bomb* simply by appealing to the fact that he is more responsible than Bianca for the threat to Clint? Or perhaps by appealing to the idea, that while they are both responsible for that threat, Al is also culpable?²⁵ If so, it looks as if there is no need to invoke the Pre-emption Principle.

The first step in addressing this objection is to stress the difference between the Pre-emption Principle and this line of argument, which is already available within the Responsibility View. In establishing who in *Bomb* is liable to be killed in order to defend Clint, defenders of the Responsibility View focus on who bears more responsibility overall for the threat to Clint. It's because Al bears greater responsibility than Bianca (or perhaps because he is not only responsible, but also culpable) that harms to him count for less in the proportionality calculation. Because of this, harming Al becomes the least harmful means to save Clint (McMahan 2009; Draper 2016).

According to the Pre-emption Principle, by contrast, what explains Al's liability is not his responsibility for the threat to Clint, but his responsibility for manipulating Bianca. Al's remedial duty is to bear the costs associated with remedying the wrong of manipulating Bianca. The two kinds of liability are different in that they have different sources. They are triggered by different wrongs.

Of course, my claim that Al is liable because of his remedial duty toward Bianca is compatible with the claim that he is also liable in virtue of the fact that he has culpably orchestrated Clint's killing. If the Responsibility View is correct, then Al is plausibly also liable to be harmed defensively for this reason, in the same way in which he would be liable for having orchestrated Clint's killing if, say, he had hired a killer, instead of manipulating Bianca. Al's liability in *Bomb* is overdetermined in this way.

Can there be cases in which the Pre-emption Principle and the traditional strategy of looking at Al's overall responsibility (or at his culpability) for the threat to V will deliver different results? The answer partly depends on how we understand the details of that strategy. McMahan, for example, believes that moral responsibility is a function of: a) how likely it is, given the evidence available to the agent, that unjust harm will occur if he acts in a certain way; and b) how serious the harm will

²⁵ The Responsibility View does not rule out appealing to considerations culpability. Its supporters do acknowledge that culpability typically is important in determining who is liable. What they object to is the claim that culpability is necessary for liability.

be, if it occurs (McMahan 2005, 397). If so, it looks as if we could imagine variations of *Bomb* in which Al and Bianca are equally responsible, since they both fulfil a) and b) in the same way, and yet Al is liable in virtue of the fact that he has manipulated Bianca.

Consider:

Bomb V: Al has apparently good, though ultimately misleading, evidence that Clint is about to detonate a bomb that will kill a large number of innocent people. Instead of trying to kill Clint himself, or persuading Bianca to do so, Al manipulates Bianca into attacking Clint. Since he worries that Bianca might not act otherwise, Al intentionally misleads her into believing that the explosion would kill someone she cares about. Bianca attacks Clint.

In *Bomb V*, we can hold constant both a) and b) for Al and Bianca. The harm Al is trying to inflict on Clint is the same as the harm Bianca is trying to inflict on him. And we can assume that, given the evidence available to each of them, the likelihood that Clint will be unjustly harmed if they act as they plan to is the same for Al and Bianca. Still, we might think that Al is liable, rather than Bianca, because he has wronged Bianca by manipulating her into attacking Clint, and he has a duty to offset the costs of the wrong she has suffered at his hands. Al would not be liable in the same way, if he had persuaded Bianca (as in *Bomb III*).

Or maybe not. Perhaps you disagree with McMahan's analysis of responsibility in terms of (a) and (b), and thus reject the claim that Al and Bianca are equally responsible in *Bomb V*. Or perhaps you don't share the intuition that Al rather than Bianca is liable in *Bomb V*. I cannot pursue these objections here. All I will say is that even if it turns out that the verdicts of my view always converge with those of the traditional formulations of the Responsibility View, this at most would show that the two are extensionally equivalent. This would be no objection to my view, because if the Pre-emption Principle is independently plausible, my view picks morally salient features of Al's liability that are ignored by any view that focuses exclusively on Al's responsibility (or culpability) for the threat posed to Clint.

5 Conclusion

We have seen that in defending the claim that liability is affected by manipulation, the challenge is to explain why being misled into ϕ -ing by manipulation makes a difference to one's liability, as compared to:

- being misled into ϕ -ing by natural events;
- being misled into ϕ -ing by someone's honest attempt to persuade us that ϕ -ing is what we have reason to do.

Why should we treat Bianca differently in *Bomb* than we do in *Bomb II* and *Bomb III*, despite the fact that she acts exactly in the same way in all three cases, by responding to the same evidence? (Bianca performs the same actions, with the same intention, the same degree of care etc.)

My view has the resources to answer this question because the Pre-emption Principle grounds Al's liability in the wrong that Bianca suffers at his hands, i.e. in the fact that Al has intentionally produced the relevant misalignment between Bianca's actual reasons for action and her apparent ones. Since there is no wrong that Bianca suffers at the hands of Al in *Bomb II* or *Bomb III*, the Pre-emption Principle does not generate liability for Al in either of these cases.

In *Bomb II*, Al is not involved at all in the production of the relevant misalignment between Bianca's actual reasons for action and her apparent ones. In *Bomb III*, Al does produce the relevant misalignment, but insofar as he does so unintentionally, Bianca is not wronged by him. In other words, in *Bomb III*, the Pre-emption Principle does not apply because clause ii) of the principle—call this the “intentionality” clause—is not fulfilled.

Here we might wonder how important the intentionality clause really is. Could we not adopt a broader principle—call it the *Pre-emption Principle**—according to which:

When B ϕ -s and the following conditions are in place:

- i. B ϕ -s in response to apparent reasons that are misaligned with her actual reasons;
- ii. the misalignment between the actual reasons and the apparent ones is produced by A's interference with B's practical deliberation;
- iii. A acts culpably or recklessly;

A, rather than B, is liable to bear (at least some of) the costs associated with B ϕ -ing, provided that doing so will secure whichever aims justify the imposition of those costs.

The difference between the Pre-emption Principle and the *Pre-emption Principle** is that the former, but not the latter, requires the misalignment between B's apparent reasons for action and her actual ones to be *intentionally* produced by A, i.e. produced *with the aim of causing such misalignment*.

Consider for example

Bomb VI: Al tells Bianca that Clint is about to detonate a bomb that will kill a large number of innocent people, knowing that she will correctly interpret this as a joke. However, Esther overhears and, not knowing that it's a joke, attempts to kill Clint to prevent him from detonating the bomb. Al has acted recklessly in telling the joke while being within Esther's earshot.

While Al would be liable in *Bomb VI* under the *Pre-emption Principle**, he is not liable under the Pre-emption Principle, since in this case Al is not aiming to produce the relevant misalignment between Esther's apparent reasons for action and her actual ones (which is why *Bomb IV* is not a case of manipulation).

Although I am somewhat inclined to accept the *Pre-emption Principle**, defending it would require further work. Notice however, that even if the *Pre-emption Principle** turns out to be true, this would not undermine the Pre-emption Principle. Far from it, if the *Pre-emption Principle** is true, the Pre-emption

Principle is true *a fortiori*.²⁶ In that case, the Pre-emption Principle would simply be an instance of a broader principle that regulates the distribution of liability in a wider range of cases, beyond those concerned with manipulation, in which someone is misled into posing a threat of unjust harm.

The reason why I have focused on the narrower principle is that my aim here was to provide an analysis of how liability is affected by manipulation. Readers who disagree about the truth of the Pre-emption Principle*, might nonetheless be persuaded by this analysis, and therefore agree about the truth of the Pre-emption Principle.

Acknowledgments An earlier draft of this paper was presented at the *Conversations on War* workshop on “War and Moral Responsibility” in Perast, Montenegro. (The workshop was 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). I’m grateful to the participants, as well as to Jonathan Gingerich for stimulating discussions. I am also grateful to Cécile Fabre, Laura Valentini and two anonymous referees (one of whom turned out to be Kaila Draper) for very helpful written comments. Finally, special thanks to Helen Frowe for discussing some of the central ideas of the paper at length.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Baron, M. (2003). Manipulativeness. *Proceedings and Addresses of the American Philosophical Association*, 77(2), 37–54.
- Bazargan, S. (2014a). Killing minimally responsible threats. *Ethics*, 125(1), 114–136.
- Bazargan, S. (2014b). Moral coercion. *Philosophers’ Imprint*, 14(11), 1–18.
- Blumenthal-Barby, J. S. (2012). Between reason and coercion: Ethically permissible influence in health care and health policy contexts. *Kennedy Institute of Ethics Journal*, 22(4), 345–366.
- Buss, S. (2005). Valuing autonomy and respecting persons: Manipulation, seduction, and the basis of moral constraints. *Ethics*, 115(2), 195–235.
- Coons, C., & Weber, M. (Eds.). (2014). *Manipulation: Theory and practice*. Oxford: Oxford University Press.
- Draper, K. (2016). *War and individual rights: The foundations of just war theory*. Oxford: Oxford University Press.
- Fabre, C. (2012). *Cosmopolitan war*. Oxford: Oxford University Press.
- Fabre, C. (2020.) *Through a glass darkly. The ethics of espionage and counter-intelligence*. Manuscript on file with the author.
- Fischer, J. M. (2017). Responsibility, autonomy, and the zygote argument. *Journal of Ethics*, 21(3), 223–238.
- Fischer, J. M., & Tognazzini, N. A. (2009). The truth about tracing. *Nouûs*, 43(3), 531–556.

²⁶ The Pre-emption Principle provides a set of sufficient, not necessary, conditions for the pre-emption of liability.

- Frowe, H. (2020). The moral irrelevance of moral coercion. *Philosophical Studies*, this issue.
- Gardner, J. (2018). *From personal life to private law*. Oxford: Oxford University Press.
- Gorin, M. (2014a). Do manipulators always threaten rationality? *American Philosophical Quarterly*, 51(1), 51–61.
- Gorin, M. (2014b). Towards a theory of interpersonal manipulation. In *Manipulation: Theory and practice* (pp. 73–97).
- Greenspan, P. (2003). The problem with manipulation. *American Philosophical Quarterly*, 40(2), 155–164.
- Hannah, J. (2015). Libertarian paternalism, manipulation, and the shaping of preferences. *Social Theory and Practice*, 41(4), 618–643.
- Haque, A. A. (2017). *Law and morality at war*. Oxford: Oxford University Press.
- Herman, B. (1993). *The practice of moral judgment*. Cambridge: Harvard University Press.
- Korsgaard, C. M. (1996). *Creating the kingdom of ends*. Cambridge: Cambridge University Press.
- Lazar, S. (2012). Necessity in self-defense and war. *Philosophy and Public Affairs*, 40(1), 3–44.
- Manne, K. (2014). Non-machiavellian manipulation and the opacity of motive. In *Manipulation: Theory and practice* (pp. 221–245).
- McMahan, J. (2005). The basis of moral liability to defensive killing. *Philosophical Issues*, 15(1), 386–405.
- McMahan, J. (2009). *Killing in war*. Oxford: Oxford University Press.
- Mele, A. R. (1995). *Autonomous agents*. Oxford: Oxford University Press.
- Mills, C. (1995). Politics and manipulation. *Social Theory and Practice*, 21(1), 99–130.
- Noggle, R. (1996). Manipulative actions: A conceptual and moral analysis. *American Philosophical Quarterly*, 33(1), 43–55.
- Otsuka, M. (1994). Killing the innocent in self-defense. *Philosophy and Public Affairs*, 23(1), 74–94.
- Parfit, D. (2011). *On what matters* (Vol. 1). Oxford: Oxford University Press.
- Pereboom, D. (2014). *Free will, agency, and meaning in life*. New York: Oxford University Press.
- Raz, J. (1999). *Practical reason and norms*. Oxford: Oxford University Press.
- Raz, J. (2011). *From normativity to responsibility*. Oxford: Oxford University Press.
- Ripstein, A. (2020). *Kant and the law of war*. Manuscript on file with the author.
- Rosen, G. (2003). Culpability and ignorance. *Proceedings of the Aristotelian Society*, 103, 61–84.
- Tadros, V. (2013). *The ends of harm*. Oxford: Oxford University Press.
- Tadros, V. (2020a). Secondary duties. In Miller P. B. and Oberdiek J. (Eds.). (2020) *Civil wrongs and justice in private law*. New York: Oxford University Press (pp. 185–207).
- Tadros, V. (2020b). Two grounds of liability. *Philosophical Studies*, this issue.
- Wood, A. W. (2014). Coercion, manipulation, exploitation. In *Manipulation: Theory and practice* (pp. 17–50).
- Yaffe, G. (2003). Indoctrination, coercion and freedom of will. *Philosophy and Phenomenological Research*, 67(2), 335–356.

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