Self-defence among Innocent People

GERHARD ØVERLAND

The Ethics Programme
University of Oslo
Norway
gerhard.overland@filosofi.uio.no

I explain the asymmetry between innocent aggressors and their (innocent) victims, and attempt to separate justified and unjustified defensive force when both parties are innocent. I propose the principle of initiating behaviour, which states that: 'In order for one person to be justified in using defensive force the other party must initiate the apparently threatening behaviour, but the defendant's interpretation of that behaviour, as being threatening, would have to be reasonable.' We can thereby maintain the view that there is a significant relation between an act being justified and which party ought eventually to be given priority in a conflict situation. If we know that an act is justified, we know that we have to give preference to its agent when there is a choice between imposing equal harm on him or her and his or her victim.

In this paper I propose a principle by which we can decide which party is justified in using defensive force when unfortunate circumstances have brought innocent parties in a deadlock in which it appears that if one innocent party does not kill the other the latter will kill the former. Typical situations may be when what appears like an aggressor is not an aggressor at all, but someone doing something making him or her look like one (the case of the apparent aggressor); or when circumstances outside a person's control has made his or her into a threat for another person (the case of an innocent threat); or when a person mistakenly has good reasons to believe that there is a justifying cause for aggression and when a person lacks the required

* Much inspiration to write this paper stems from reading chapter 6 of Arthur Ripstein's book Equality, Responsibility, and the Law (Cambridge: Cambridge University Press, 1999), and I am grateful for his encouraging words. An earlier version of this paper was presented at the 21st IVR (Internationale Vereinigung für Rechts- und Sozialphilosophie) World Congress in Lund, Sweden, and later at a seminar held by the Ethics Programme, University of Oslo. I am indebted to Antony Duff for helpful comments on an earlier version of this paper. I would also like to thank Kai Dramer and John Richard Sageng for stimulating discussions, and Jakob Elster for written comments.

mental capacity to evaluate his or her actions in order to be held morally responsible for them (cases of an innocent aggressor).

While we may believe the grounds for acting of the typical innocent aggressor to be reasonable, it might also be the case that the innocent aggressor lacks the capacity to evaluate his or her grounds for acting. In neither case could the agent be held morally responsible for his or her actions. An innocent threat is a person who is causally involved in a threat of harm to another person but not through his or her own agency. An innocent threat is not doing anything; it is merely something that happens to the person that makes him or her into a threat. Apparent aggressors are people who look like aggressors but who in fact are not engaged in aggressive action. Hence a person using defensive force against an apparent aggressor may be an example of an innocent aggressor.

Although it is widely accepted that use of defensive force is justified in cases when the aggressor is a culpable person with an intention and ability to kill, what its minimal requirements should be is contested. My proposal is that in order for one innocent person to be justified in using defensive force against another innocent person, this other person must have initiated the apparent threatening behaviour, but the defendant's interpretation of that behaviour as being threatening must be reasonable. This view implies that although reasonable beliefs do not always justify, a reasonable interpretation of an initial behaviour by another party as being threatening does.

Preliminaries

It is common to differentiate between two main usages of the term 'justification'. In one version, 'justification' is used to designate an objective and impartial perspective, which implies that a person is justified when he or she acts in accordance with what it would have been permissible to do. The circumstances that make an agent's act permissible are the same circumstances that make it a candidate for being justified. However, in order to be justified the agent must know about those circumstantial factors. This observation implies that all permissible acts need not be justified simply because the agent may not know about the justifying circumstances, but whenever the agent does, and only then, will he or she be justified in doing so and so. One attraction of such a true belief view of justification is that saying that a conduct is

3. One could, of course, include all permissible action in those that are justified, and thereby deny the knowledge requirement. Such a pure objectivist idea of justification would be coextensive with what is permissible, but is not, I think, a plausible candidate for how we should use the term.
justified has implications for what we take to follow from it, namely that the conduct may be supported and not prevented.4

In a second usage, what determines whether or not an action is justified is evaluated from the perspective of the agent. According to this view a person might be justified in acting in a particular manner if he or she acts on his or her beliefs about a situation and those beliefs are reasonable. One attraction of such a reasonable belief view of justification is that it matches up well with instances in which we think it appropriate to exempt a person from moral blame. One of the difficulties of this view, however, is that it must deny the existence of an important relationship between saying that an action is justified and whether or not it may be prevented. Moreover, it may very well allow for two parties being justified in using defensive force against each other.

Confronted with a moral wrong, we can assess it with regard to two parameters. We might think it is permissible to prevent the wrong, and/or we might think that it is appropriate to blame the person behaving wrongly. To say that an agent is doing something morally wrong is either to give an assessment of the external circumstances the agent is bringing about or to assess the internal relations of the agent’s intentions, desires, and beliefs, or a combination of the two. The external has to do with the question of permissibility, while the internal concerns questions of blame. It is important to note that blame-worthiness and permissibility should be kept apart, and that a moral wrong could concern either of the two, or both.

Granted this taxonomy, a justification is easily placed on either of the two sides of the morally wrong. The reasonable belief view of justification accentuates exemption from blame, while the true belief view of justification accentuates permissibility of prevention. I want to propose a third way of using the term ‘justification’ based on who might be prevented in situations of conflict. While I shall give up the requirement of true beliefs, I retain the view that if an agent is justified in doing so and so, it should have implications concerning who ultimately might be prevented from acting in such and such a way when party to a conflict. Yet the relation between an act being justified and the question of who might be prevented is more complicated than has normally been assumed. My proposal is that when there is a question of imposing equal harm on either of two parties, the one who is justified ought to be given priority and not be prevented. Accordingly, when a person is justified in using defensive force, it is impermissible to impose any amount of harm on the defendant in order to save the (apparent) aggressor or threat from an equal amount of harm.

It may pay to mention that all instances in which adherents of the true belief view on justification would say that an agent is justified in doing so and so, the agent would be justified according to my proposal as well. However, my proposal also covers some, though not all, instances in which adherents of the reasonable belief view would say that an agent is justified in doing so and so.

1. Apparent Aggressors

Let me begin by introducing an example in which it is quite reasonable that a person uses defensive force, but nevertheless permissible to prevent it. Imagine that one person attacks another person, firing a loaded gun. Unbeknown to either of them, the remaining bullets in the chamber are flawed, and will not fire. The aggressor in this example is a culpable apparent aggressor, which means that we are not speaking of self-defence among innocents. The aggressor has an unjustified intention to kill his victim, and takes what he thinks are the necessary steps towards doing so. By chance, those steps are inadequate.

Clearly, such an attack must pass the test of reasonable fear, even if chance events mean that the aggressor no longer poses any mortal danger. Many people will say that use of defensive force in this case should be deemed justified. The epistemic situation for the defendant is, after all, the same with regard to this apparent aggressor as it would have been with regard to a real one.

If we assume that a justified act may be supported but not prevented, we cannot grant the victim of the apparent aggressor a justification. Since the last bullets in the aggressor’s gun were flawed, it could be permissible to prevent the defendant from killing him, for example, by snapping the gun out of his hands. She is not under any real threat, and any use of defensive force from her side would be force used in vain. If we therefore want to say that a person may be justified in using defensive force against an apparent aggressor, it looks as if we must give up the idea that there is any interesting relation between an act being justified and whether or not it might be prevented. Given the fact—if it is a fact—that it would be permissible to prevent the person under attack from defending herself, it appears to be of little interest to say that use of defensive force would be justified; nothing seems to follow from it. If all we want is to exempt from blame, an appropriate moral excuse would do.

We might observe, however, that under no circumstances can the upshot be that it is the culpable aggressor, although only apparently representing a threat, who should be selected when there is a choice of saving the aggressor or the victim. Although the victim might be prevented from killing the aggressor, there are limits to the means that may be used in doing so. To inflict considerable harm to the prospective victim in order to save the culpable apparent aggressor is not permissible.

Innocent Apparent Aggressors

There are situations in which people have good reasons to regard a person as an aggressor and to take appropriate steps against him or her, but where the person mistaken for an aggressor has no intention to harm. A person who merely looks like an aggressor is an innocent apparent aggressor, and should be distinguished from a culpable version who intends to harm but who, unbeknown to his or her victim, does not possess the necessary means to do so.

I shall now discuss the plausibility of saying that some victims of innocent apparent aggressors may be justified in using defensive force, and what the criterion for it could be. As I said, I will give up the requirement of true beliefs, but maintain the view that if an agent is justified in doing so and so, it should have implications for the question about who ultimately might be prevented when there is a conflict between two parties. I will therefore use this insight when evaluating the different proposals.

Criterion 1: Reasonable beliefs on part of the defendant. It has been argued that the use of defensive force based on mistakes of fact, where the defendant has reasonable beliefs concerning those facts, is justified. And clearly, one reason for claiming that a reasonable mistake is justified is to differentiate it from a simple excuse based on incapacity on the part of the defendant.

When a mistake is reasonable, a person acts intentionally on reasons she takes herself to have, and we can see that those reasons really were reasonable in her situation. She has demonstrated due care, and acted upon the reasons she has been able to gather in a way we want people to do. There is nothing more we should expect from her. Therefore, we could be tempted to say that her act is justified.

My proposal is that if reasonable beliefs justify, and we have to choose between imposing equal harm on either of two parties in a heated situation, we would have to give preference to the one acting upon reasonable beliefs. In talking of an escalating situation, I imagine a person perceiving herself to be under attack and therefore setting out to defend herself. Subsequently, the apparent aggressor defends himself with available means against the defendant (assuming that he carries a gun with him, for example). In such an escalated scenario, we have two people fighting each other with lethal means.

At a minimum, we can now see that talk of what is reasonable is deficient. There might be situations in which a person reasonably believes herself to be under attack, but where it would be impermissible to favour the defendant if we had to choose between imposing equal harm on either of the two. Let us assume that Mary believes she must defend herself against Bill with lethal force. She believes so because a third party has told her untrue things about him. She has been told that Bill is planning to kill her, and that he is going to do so in such and such a way. But what she has been told is false; Bill is not intending to kill anyone. Let us furthermore assume that what Bill does would not in ordinary circumstances have been enough to make Mary believe that he was an aggressor. The reason why she believes so in this case is what
she has been told about him, and that he (accidentally) acts in a way that fits well with that description. Nor has she cause to suspect that what she has been told is not true. The latter is what makes her beliefs reasonable, provided, of course, that Mary has demonstrated due care, and that there is nothing more we could expect from people in such a situation. I shall call this example 'OriginalMary'.

In OriginalMary, Mary acts on the basis of reasonable beliefs when she begins to use defensive force against Bill who is acting in a manner consistent with the warnings given to Mary. Yet if it comes to a choice between letting Mary kill Bill, or letting Bill kill Mary, it seems wrong not to let Bill kill Mary. This creates a problem for the reasonable belief view of justification. It seems odd to claim that Mary is justified when we would judge it wrong to save her at the cost of the life of the one she is defending herself against. We should not, I believe, on the one hand, say that Mary is justified in using defensive force against Bill, while we, on the other hand, perceive that if it came to a choice between them we would have to assist Bill. This observation indicates that acting on reasonable beliefs is not enough to grant a person a justification, nor to guarantee that a person doing so will be favoured if there is a choice between the two parties.

If one holds the view that Mary as Bill equally merit being saved, one could maintain that both are justified in using defensive force: Mary would be justified in using defensive force against Bill's apparent attack, and Bill would be justified in using force to defend himself against Mary. This proposal is not very illuminating because what more would we in this case say that neither of them should be blamed for using defensive force. It would therefore be better to say that neither is justified in using defensive force, and merely grant each an excuse. Yet, I should think that many people would regard it permissible to save Bill at the cost of Mary, in which case we had better say that he is justified in using defensive force, while she can only be excused.

Although it might be reasonable to regard a person as an aggressor if a third party has provided untrue information about him or her, you would not be the one deserving of leniency in a situation in which we have a choice between imposing equal harm on you or the innocent apparent aggressor in order to save one person. It is after all, you that triggered the escalating process.

Criterion 2: Blame on part of the apparent aggressor. An innocent apparent aggressor may or may not be to blame for appearing aggressive. It would depend on what he has done preceding the situation in which someone believes herself to be under attack from him. Perhaps, then, when deciding whether or not a person is justified in using defensive force, what is essential is whether or not the apparent aggressor has done something he should have known could cause other people to regard him as an aggressor?

6. It is not enough that we could not expect more from Mary in particular; she must live up to the requirement of a reasonable person.
This proposal means that we add something to the reasonable belief view of justification. In order to be justified one must act on reasonable beliefs and it must be the case that the person against whom defensive force is used is to be blamed for appearing like an aggressor. If the apparent aggressor ought to have known that his behaviour would reasonably be perceived as so threatening by other people that they would think use of defensive force were necessary, and he could have avoided such behaviour, he is to be blamed for creating that impression. Accordingly, I shall call such a person 'a blameful apparent aggressor'. Be aware that he is not blameworthy because he is a culpable aggressor with an intention to harm: the blame is only concerned with why he appears like an aggressor.

The proposal at this point is that if response to a blameful behaviour by the other party can be justified on the basis of reasonable belief, and we have to choose between harming either of the two parties equally in an escalated situation, we would have to choose not to harm the one responding to the blameful behaviour. Thus, a person may be justified in using defensive force only if the apparent aggressor is to be blamed for creating an aggressive impression; he must have done something he ought to have known was menacing. Put otherwise, he must at least be partly responsible for the situation, and more so than the other party. And if none of the parties are to be blamed for the unfortunate escalated situation, none of them will be justified in using defensive force either.

Now, it may seem correct to give the defendant moral precedence if the apparent aggressor is to be blamed for appearing like an aggressor. This becomes obvious if we imagine Bill is simply having fun with Mary. If Bill deliberately creates a situation in which he appears like an aggressor, it would be unfair to let someone else carry the cost of the situation created. Moreover, if Bill behaves recklessly or negligently he ought to carry the cost of doing so.

That said, this criterion and the one on reasonable beliefs on the part of the defendant share a significant problem. They cannot help us to resolve all instances of escalated force. The above example illustrates the problem. In OriginalMary, Mary believes she must kill Bill because a third party has told her untrue things about him, not because of something he has done. Bill is therefore not responsible for appearing to Mary to be an aggressor. But Mary, on her part, has every reason to believe what the third person has told her about Bill, and she is therefore not to be blamed for believing that she must defend herself against him. Thus, in OriginalMary it is not only the case that Mary acts on reasonable beliefs, but it is also the case that neither Bill nor Mary are to be blamed for the situation.7

7. Circumstances in which this may happen may readily occur in wartime. Soldiers may have been led to believe (falsely) that their attack is justified.

In OriginalMary, the only person to be blamed is the informer who told Mary untrue things about Bill. But even this person could have been telling untrue things about Bill in good faith.
Notwithstanding the fact that neither is to blame, it seems wrong not to choose to save Bill if the situation escalates and we have to choose between saving him or Mary. This indicates that Bill is justified in using defensive force against Mary, although Mary is not to blame for escalating the apparent conflict. We can therefore not say that the idea of being blamed for causing an impression of impending aggression is sufficient to narrow down the reasonable belief view of justification and thereby determine when we should say that a person is justified in using defensive force. Often, however, the person who is morally responsible for appearing like an aggressor must carry the cost if the situation escalates. But there is another criterion, which in these cases coexists as an explanation.

Criterion 3: Initiating behaviour. In the example of Bill and Mary, both may be acting upon reasonable beliefs and neither of them may be to blame for appearing like an aggressor. It seems nevertheless that there is a reason for giving Bill priority if there is a choice of inflicting equal harm on one of them to save the other. Why?

The simple answer, I believe, is that Mary initiates the escalating process. She starts to do something, which triggers Bill's response. Because although Bill appears like an aggressor to Mary, Bill is not doing anything that would have been regarded as threatening if it were not for the particular information given to Mary from a third person.\(^8\)

The idea to be investigated at this juncture is that we should always give preference to the one who does not initiate the escalating process. If it makes sense to say that one person's action is a response to something the other person does, the first person, who begins the process, will have to be sacrificed if it comes to a question of saving one of them. The one justified in using defensive force is the one responding to an apparently threatening behaviour by the other party.

One problem, of course, is that a person's response might be triggered by a variety of irrelevant factors. For example, if I should regard your tapping on the desk as an attempt to blow up the building, my reaction would not be deemed justified if I attacked you with all my force. And this is true although I respond to something you do, namely your tapping on the table.\(^9\)

It seems as if a requirement of reasonableness must be called upon again. Any interpretation will not do. The defendant's interpretation of the behaviour

---

8. While one might be justified in blaming the informer for causing an escalation of the conflict, i.e., identify her as the initiating party, it would nevertheless be correct to say that, relative to Bill and the apparent aggressor (Mary), Mary was the initiating party. In situations under discussion, it is never the case that the defendant may survive by using force against third parties. If that were the case, it could make sense to discuss the permissibility of using defensive force against such third parties. I shall not do so here.

9. What if it turns out that your tapping on the desk would have blown up the building? In that case the interpretation of your tapping on the desk as threatening would be unreasonable, but true.
of the apparent aggressor as threatening has to be reasonable; if not, the response cannot be justified. The interpretation of the epistemic situation must be reasonable because we need to know that it was the apparent aggressor who initiated the escalating situation and not a particular belief on part of the defendant. Thus, whether or not Mary is justified in killing Bill, or Bill is justified in killing Mary, depends (1) on which of them initiated the escalating process, and (2) whether interpreting such initiating behaviour as threatening is reasonable.\(^1\)

At this point I propose the principle of initiating behaviour, which reads as follows:

In order for one person to be justified in using defensive force the other party must initiate the apparently threatening behaviour, but the defendant's interpretation of that behaviour, as being threatening, would have to be reasonable.

The principle of initiating behaviour explains the asymmetry between innocent aggressors and their (innocent) victims, and the principle separates justified and unjustified defensive force when both parties are innocent.

The principle of initiating behaviour narrows down the reasonable belief view of justification further than the criterion of blame. On many occasions, however, they will be coextensive. By accepting the principle of initiating behaviour as pertinent to identify justified agents, we may maintain the view that if an agent is justified in doing so and so, it will have implications for determining who ultimately might be prevented when two parties are locked in conflict. The implication is that when there is a question of imposing equal harm on either of two parties, the one who is justified ought to be given priority.

2. On Initiating Behaviour

The person to be sacrificed does not need to do something he or she ought to have known involved a risk.\(^1\) The only thing that is required is that it makes sense to say that what initially triggered the escalation can be traced to one of the two parties. Whenever that is possible, it determines which party should be deemed justified in using defensive force among innocent people.

10. The concept of the reasonable is not always transparent. Let me first say that my use is not normative in the sense 'the reasonable' is used in political theory. However, it is not purely descriptive either. An interpretation of some conduct as threatening will depend upon the norms of the particular social culture, and the real level of risk one is exposed to in that culture. As has been indicated by colleagues here in Oslo, what may reasonably be seen as threatening on the streets of New York may not always reasonably be interpreted as threatening on the streets of Oslo. It is beyond the scope of this paper, however, to attempt to make the epistemic concept of the reasonable transparent.

11. Thus, I disagree with Ripstein, who says that in doing something that might reasonably be perceived as an attack, 'I take a risk that I will be killed or wounded by the reasonable defensive measures of another' (Ripstein, Equality, Responsibility, and the Law, p. 195).
Tracking the Initiating Behaviour

I said that interpreting an epistemic situation as threatening would have to be reasonable because we need to know whether it was the apparent aggressor who initiated the escalating situation and not particular beliefs on the part of the defendant. However, retracing such initiating behaviour is not always a straightforward process.

To begin with we could say that the requirement of a reasonable interpretation implies that most people who knew nothing about the apparent aggressor would regard him as an aggressor by simply observing the movements of his body and hearing what he says. Yet this will not always do because we may know something about the person that we would like to take into account when interpreting his behaviour. We therefore need to know whether taking such knowledge into account helps us trace the fount of the initiating behaviour.

Should, for example, things done in the past count as initiating behaviour? Imagine a person with a record of a contract killer, but who recently converted to a more peaceable occupation. You meet him, and he acts in such a way to warrant a reasonable expectation on your part of danger, especially as you believe he still pursued a career as a contract killer. You know about his record, but not his recent transformation, and reasonably believe that he is on a contract to kill you. Under such circumstances, if the conflict should escalate, and we have to choose between saving you or the assumed contract killer, it would be wrong according to the principle of initiating behaviour to save the latter. We would have to act to save you because your beliefs about your supposed opponent rest on his former proven conduct and can therefore be traced back to him. The important thing is not the exact timing, but whether an impression of danger flows from the person due to his past or present behaviour.

What about initiating behaviour that will make someone believe something about an initiator that is not true? Imagine that a person A has behaved in the past in such a manner as to persuade others that A’s present behaviour is dangerous in some way or another. In this case, too, priority should be given to the one who responds to A’s present behaviour, provided that interpreting A’s behaviour as threatening is reasonable in light of his (false) record of aggressive behaviour. Mary may, for example, be justified in using defensive force against Bill if it were reasonable for her to interpret Bill’s behaviour as threatening in light of (false) information she possesses on his putative criminal record, provided that something about Bill’s past behaviour made it reasonable to interpret him as having such a record. The difference between this version of the example and OriginalMary is that the reason for the mistake flows from Bill and not from a third party who tells Mary untrue things about Bill. It is Bill himself whose past behaviour warrants an interpretation of his present behaviour as threatening. In general, we can say that when tracing initiating behaviour it is important that the information, in light of which an interpretation of a particular behaviour as threatening can be said to be reasonable or not, in some way or another flows from the agent in question.
Particularly hard cases seem to come about concerning mistakes of identity. In such cases we might hesitate before determining the initiator. One problem is that it would be difficult to decide when it is common knowledge that looking a particular way may be dangerous. This has to do with the criterion of reasonableness. Imagine, for example, a person who grows a beard, which makes him resemble a well-known dangerous person. Given the fact that he now resembles such a person, it might be reasonable to interpret his behaviour as threatening. But should the growing of a beard count as the initially threatening behaviour? To me it seems as difficult to say that the beard grower is the initiator, as it is to determine who should be saved at the cost of the other. It is not clear that what makes this person appear like an aggressor actually flows from him, or whether changes in the world has made him appear like an aggressor. In certain circumstances it might prove hard to determine the initiating party, and in those cases it should come as no surprise that we do not know how to settle the question.

The principle of initiating behaviour can allow for border cases like the beard case, but will not sit well with examples in which it is clear that the party who initiates the dangerous situation is, at the same time, not the person who has to carry the costs if the situation deteriorated. The latter would be a counterexample to the principle. No such example has yet been identified. Border cases are different; they are cases the theory is not apt to settle. Yet, since we have no clear intuition about these cases, they do not indicate a failure with the theory.

Questioning the Action Requirement

So far, I have taken for granted that in order to be justified in using defensive force, the defendant must respond to something the other party does, or has done in the past. In spite of this, it is not clear that in order to be justified in defending oneself against an innocent person, the threat has to be constituted by a person doing a particular action.

I have already given up the requirement that a person ought to have known better when he or she acted as s/he did. In OriginalMary it was not the case that Mary was to blame for using defensive force against Bill, yet she would be the one who had to carry the costs if the situation escalated. And having accepted that the person who eventually has to bear the costs does not need to be morally responsible for causing the conflict to escalate, why should we insist that the initiating behaviour must be an action at all? If we accept that a person might be held responsible (in the sense that he or she eventually will have to carry the costs) for things he or she is not to be blamed for doing, and perhaps may not even know that he or she is doing, then there should be no reason to resist the assumption that a person might be held responsible in the same way for things he or she does not do either, as long as we can say that it is what happens with him or her that initiated

12. Douglas N. Husak presented this tricky example to me in conversation.
the apparently dangerous situation. There seems to be no reason to require agency if we have given up the requirement of intentional action.

Let me introduce an example with an innocent threat. Although it has been contested, I shall assume the following to be an instance of justified self-defence.\textsuperscript{13} Imagine Lisa walking in a park, and a sudden gust of wind blows her down a well. It would be permissible for a person at the bottom of it to disintegrate her falling body by using his ray gun,\textsuperscript{14} provided, of course, that he needed to do so to save himself from being crushed by Lisa's falling body. We assume that Lisa will survive if not disintegrated, because the person already in the well acts as a protective buffer from the lethal impact.

Now, if we accept that the person in the well is justified to use defensive force, what may be the reason for it? Remember that both are morally innocent with regard to the unfortunate situation they suddenly find themselves in. It makes therefore no sense to say that defensive force is justified because Lisa ought to have known that she might unwittingly threaten the life of another. A reason at hand is to say that the person in the well is justified in using defensive force because it is the falling person who initiated the dangerous situation. Lisa is a threat to the life of the person down the well, not the other way around. Although Lisa is not doing something that constitutes threatening behaviour, it seems plausible to say that she is the initiator of the situation.\textsuperscript{15}

By saying that Lisa initiates the dangerous situation, we also say which of the two people should carry the cost of it. It is therefore not permissible for Lisa to save herself at the expense of the other person. This observation does not necessarily imply that we would blame her if she did. There may be reasons for granting her some form of (hardship) excuse.\textsuperscript{16} However, Lisa would


\textsuperscript{14} If you do not like the idea of a ray gun, imagine that the person in the well can deflect the falling person further down the shaft.

\textsuperscript{15} One might question if this is a correct use of the word 'initiate'. Normally it means something like 'to begin' or 'set going', which are actions, and not something that happens to a person. However, the point I am stressing is that to initiate is to cause something, for example, an instance of violence. If two kids are fighting, we might ask who started the fight. And although the incriminated child only happened to trip and fall towards the other, we might agree that he nevertheless initiated the fight, but that he is not to be blamed for it. One might reply that it is nevertheless wrong of the other child to respond by hitting back. Well, that is true if he hits back after the first event is over, and merely to get revenge. The point, however, is that it is not wrong for the child to act to prevent being hit by the falling body.

\textsuperscript{16} A hardship excuse is an excuse we may grant people who do something wrong, but for which we find it difficult to blame them because they are in a pressed situation. Other instances where hardship excuses may be relevant are where a person pushes another person
at least be a candidate for blame if she took steps to defend herself from the person in the well (and not, of course, for being a threat in the first place). For Lisa to prevent the person in the well from defending himself is wrong.

What I have said seems to be true as well if the falling person is only an apparent threat. Imagine that, unbeknown to either of the two persons, a net has been strung across the well that will catch Lisa before she hits the person at the bottom. Accordingly, when the person in the well believes only his ray gun will save him, he is mistaken. He does not need to do anything to survive. To create the sort of dilemma we are discussing, the situation has to escalate and we must imagine Lisa being able to draw her gun while falling to defend herself against the ray-gun-firing person below. In such an escalated situation it would be impermissible to save Lisa at the cost of the person in the well because it is her bodily movements that set the situation on its escalatory course. And this is true, although it is not Lisa who moves her body, but wind and gravity.

3. The Moral Significance of Initiating Behaviour

How do we explain the moral significance of being an initiator, when all it takes to become one is that you are taken by the wind? One reason why Lisa’s initiating behaviour makes a moral difference could be presented by appealing to the notion of having one’s rights violated. The assumption in this case would be that one could only be justified in using defensive force against another person if this other person had lost his or her right not to be harmed. We would then have to show that the initiator (Lisa) violated some rights of the defendant, which made the use of force permissible. A problem with this account is the difficulty it introduces in making sense of the idea that being blown down the well involves the violation of some of the rights of the person at the bottom. It might seem odd to say that a person violates another person’s rights merely because something happens to her; moreover, it is not clear what sort of rights may be violated by simply appearing to be an aggressor.

Those questioning the assumption that it is permissible to use defensive force against innocent aggressors and threats could ground their argument on the claim that any person has a right not to be (unjustly) killed, and that one cannot lose this right by accidental events in the world. In order to lose one’s right not to be killed, one has to be morally responsible for a particular behaviour, which causes another person to be under threat. According to David Rodin, for example, ‘to take the language of rights seriously, is to be committed to the idea that a person’s right may be infringed or forfeited only on the basis of something that the person is or does as a moral subject’.

---

I am not sure if we should take the language of rights that seriously. Clearly, rights may be altered by the right-holder's own deliberate actions, but they may be altered by what happens in the world as well. When you kick my leg in order to fulfil the whims of a villain who is threatening to kill you if you don't, it is not the case that I have violated any of your rights upfront. Nevertheless, it seems that you would be justified in kicking my leg. Even though I, under normal circumstances, have a right not to be kicked in the leg by anyone, I may lose this right if people around me broach danger unless I get kicked in the leg. Perhaps you would say that this is true for trivial rights, but not about the right of not being killed, for example. But, then, what should we say about the person on a track or a road towards whom we steer a trolley or car in order to save five other people standing in our way? Most people would think such acts are permissible, but the sacrificed person has clearly not violated anyone's rights.

Perhaps we could say that people have a right to use defensive force against anything that comes their way, be it a stone, trolley or a person. Such threats may or may not have violated the victim's rights. Any person, who initiates threatening action but who defends him or herself against a victim, violates the victim's right to use defensive force. That is why threats like that posed by Lisa are liable for moral blame if they act to prevent others from defending themselves against them. They are not to be blamed for what has happened to them, causing them to initiate threatening conduct, but they might be if they try to extricate themselves from their problems by shoving them onto others.

Unfortunately, this response will not take us very far. Saying that people have a right to use defensive force against anything that comes their way would delegate the same kind of right to the person initiating the threat. The initiator would have a right to use defensive force against her victim when he uses defensive force against her. There must be some account of the asymmetry between the two parties. A plausible theory must be able to explain why one party is justified in using defensive force, while the other is not. The reason for the asymmetry provided here is that it is only the person who defends him or herself against an initiating (apparent) threat or aggressor that is justified in using defensive force. But what we are looking for at this juncture is a reason that can explain the moral significance of being the initiating threat in the first place.

18. Or we might simply say that rights come with a set of almost unlimited qualifications.
A Contractual Perspective
According to Jeff McMahan, who is looking for a justification for killing innocent aggressors but sceptical about the prospects of finding any, it might have considerable social utility to permit self-defence against any threatening person other than just aggressors because one can seldom be certain that an aggressor is indeed fully innocent. And since cases involving innocent aggressors are rare, a case in which there is uncertainty is more likely to involve a culpable aggressor than an innocent one.\(^\text{20}\) Having said this, McMahan expresses a worry about morality being merely conventionalist. But the point I am about to explore is that people have good reasons to accept the principle of initiating behaviour, and not the mere fact that they have converged towards it; there are contractual grounds for accepting initiating behaviour as a demarcation principle.

Any rule that allows the killing of innocent people must be given a rationale that people in general would regard as a good reason for accepting the rule. Giving such a rationale is the essence of giving a contractual account of the rule. The point is neither that people have made a particular contract, nor that we may be confident that they would if they had an opportunity to do so. The point is that people have good reasons to accept the rules. We must focus on what we know about them in virtue of which we have reason to believe that they would have consented to the rule.\(^\text{21}\) And I should think that we know at least two things about them: (1) They think that it is a bad thing that people are killed, and (2) they think that it is worse that innocent people are killed than culpable.

Knowing this, there seem to be good reasons for accepting a rule that gives priority to a party that responds to an initiating behaviour. First, it may work as an incentive to prevent the causes of such unfortunate situations. Secondly, being an initiator will often be a proxy for culpability, and by giving priority to the responding party more innocents will be saved in the long run. Thirdly, giving priority to the responding party seems to be the best way to avoid situations in which a culpable person survives at the cost of an innocent person.

The Incentive Argument
Saying that it is always the initiating person who eventually has to carry the cost of an escalating conflict could be the best way to minimize the likelihood that such situations arise in the first place. People will try to avoid becoming initiators when they know that, whether or not they are morally responsible for so doing, they will have to carry the cost of it.

It is not always clear whether or not someone could have avoided a certain situation; innocence comes in degrees. This is true of many kinds of innocent

---


aggressors, for example, people who lack the mental capacity required to be held morally responsible for their actions. It is far from clear at which point they should be considered completely outside the realm of moral responsibility, and although we might think it is inappropriate to hold them responsible for their actions in the sense that they are punishable, we might think that it is better if they carry the costs rather than their victims in situations of self-defence.

One problem with the incentive explanation is the point that people cannot take steps to avoid certain situations when they do not know how to avoid them. After all, the crux of the example of Lisa falling down the well was exactly that she could not have known that she would be thrown off balance by a gust of wind. This observation appears to obstruct the prospect of explaining on contractual grounds why we should give priority to a person who responds to an initiating behaviour by another party, when what caused the initiating behaviour were merely something that happened to the person.

Such objections can be met, however. First, we could ask how the person down in the well possibly could know whether Lisa had demonstrated appropriate care. On what criterion could we decide whether he is morally justified to defend himself against the falling woman? Whether or not Lisa is to be blamed for falling down the well is difficult to decide both for prospective victims and bystanders alike. We therefore need an account that helps us converge on the party who ought to be given priority; both the defendant and intervening bystanders need external features which can enable us to pass judgement. Using the initiating behaviour as a demarcation principle provides such an external feature.

Moreover, although the point of the example with Lisa was that she could not have foreseen that she would be bowled over by the wind, deciding whether or not she ought to have known is not easy. Although Lisa is not to blame for choosing to walk in a particular direction at a particular time of day, it seems easier for her to avoid becoming a threat than for the person down in the well to foresee that she will come hurtling down towards him. The decisions that made up the dangerous situation are closer to her. It seems therefore rational to put the burdens on the person who creates the costs, and not on the one whom the costs fall. This observation helps explain why people's responsibility is to avoid becoming a threat to another person, rather than to avoid the use of force against initiating threats with no culpability.

The Proxy Argument
Being an initiator will often be a proxy for culpability. People who are morally responsible for endangering other people will normally also be the initiating party. And since being an initiator is an external feature that it is possible for other people to observe, sacrificing the initiator is likely to save more innocent people in the long run while putting paid the culpable. That seems to be a good reason for accepting initiating behaviour as a demarcation principle.
Clearly, it will be crucial to explore if there could be upshots from sacrificing the initiator in the form of costs to the innocent rather than the culpable. This would concern conflicts involving an intention to harm by the responding party, and the innocence of the initiator of the threatening situation. Although we might with some ingenuity imagine such situations, they will be rare. To illustrate we may adjust the example of Lisa falling down the well. Imagine that Lisa’s fall would flatten somebody intent on killing her anyway. Evil sits deep down in an empty well waiting for Lisa to come looking for water. Evil has the intention of shooting her as soon as he sees Lisa peer over the rim. But because his concentration fails him, he fails to notice Lisa who, in the darkness, is suddenly plummeting towards him at full speed. Lisa slipped on the wet ground, missed her balance and fell into the well. Evil thinks (correctly) that the only way he can survive is to use his ray gun and disintegrate Lisa before she hits him.22

People I have acquainted with this case say that it would be permissible to save Lisa at the cost of Evil defending himself. This they continue to believe, although they maintain that in situations in which both parties are innocent, priority should be given to the party in the well who is responding to the threatening behaviour of the falling person. Thus, although initiating behaviour normally sways the balance, it seems as if culpability prevails whenever it comes into play. But this should not surprise us. We comply with the principle of initiating behaviour in order to reduce the likelihood of saving culpable people at the cost of innocent; we should therefore not comply with the principle when we know that one party is culpable. If we do not know which of the two is culpable, the principle of initiating behaviour will dictate us to save Evil because intentions are not external while being the initiating party is. That is unfortunate, but, as I said, such instances will be rare.

**Fairness**
An important observation is that people in general have reasons not to accept rules which would save culpable people at the cost of innocent. We know that it may be difficult for a person under attack to decide whether or not an aggressor or threat is innocent. A rule which required the defendant to make sure that the aggressor/threat was culpable before defensive force could be deemed permissible, would lead people to give up their life on wrong occasions. Wrong occasions would be instances in which the defendant may mistakenly believe a culpable aggressor to be innocent, and therefore infer that he ought not to defend himself. In such cases a culpable person would survive at the cost of an innocent person. This may look unfair.

To avoid such unfortunate results, one could say that whenever the defendant knows him or herself to be innocent—a question easier to settle for the defendant than the question of the aggressor’s innocence—the defendant will

---

22. There might or might not be a net across the well that will catch falling Lisa before she hits Evil.
not be blamed if he or she uses defensive force against apparent threats or aggressors. People in general have good reason to allow a rule permitting self-defence against initiators because compliance to the rule does not (normally) lead to the deaths of the innocent while saving the culpable. On the contrary, such action will normally save the innocent at the cost of the culpable. Giving priority to the person who responds to an initiating behaviour will enable people to avoid giving up their life on wrong occasions.

In Original Mary, where Mary has been told untrue things about Bill, Mary initiates the escalating situation, but she does not know that by doing so she inadvertently makes it permissible to kill her if need be. Thus, by allowing a person who responds to an initiating party to use defensive force, innocent people will on occasion be killed. But that would have been the upshot whatever rule were chosen, unless we followed a rule prohibiting the use of any defensive force, or permitted it only when it was known that it was used against a culpable aggressor.

On some occasions, a rule prohibiting any use of defensive force would save people from killing an apparent aggressor. This would be a pacifist rule. It is questionable, however, whether such a rule would cause fewer deaths in the long run. A more modest version would say that it is permissible to use defensive force if and only if the aggressor is a culpable real aggressor. This would be the proposal presented by Michael Otsuka, and recently supported by David Rodin; namely that it is only permissible to use defensive force against a person who is morally responsible for being an aggressor or threat. An attractive feature about this proposal is that it points to a property about which it is easy to agree has moral significance, namely moral responsibility. However, it fails with regard to a variety of examples, such as those discussed in this paper. Neither would it, I believe, be the preferred solution on contractual grounds. This is because it will lead to plenty of situations in which a culpable person is saved at the cost of an innocent person.

Concluding Remarks

Someone comes at you, and to all intents and purposes he appears to be intent on killing you. Perhaps he is screaming blue murder, perhaps you are in a war zone and he is wearing a uniform like those who killed your friends. What should you do? Well, first of all you should consider whether he represents a real threat, and not simply someone appearing like one. The standard of evidence you have to satisfy in order to determine whether the situation actually is dangerous is the standard of reasonableness. Secondly, you have a duty to consider necessity and proportionality. Any use of defensive force should be the least available, and be proportional to the threat. If these requirements are satisfied, that is, you reasonably perceive that the threat is

real and assume you cannot escape unless you kill the aggressor, you may do so. If it is reasonable to interpret the behaviour of the other party as threatening, you are justified in using defensive force even though it should turn out that the threat is only an apparent one. But even then, we might prevent you from wounding the aggressor if we know what you don’t, i.e. that the threat is only apparent, and that we can defuse the situation at moderate costs to you. The fact that you are justified only dictates that we ought to give priority to you if there is a choice of imposing equal harm on either of the two in order to halt the escalating process.

So far there may be little disagreement. The real controversy is about culpability. I claim that although you should be able to figure out that the aggressor is not morally responsible for his or her actions, the implications of that information would be limited. Imagine that you receive a phone call telling you that the person attacking you is insane, sleepwalking, a soldier believing falsely that he is fighting a just war, or something else removing his moral responsibility. It would nevertheless not be the case that you ought to stop defending yourself, and allow the aggressor to have his or her way. If it comes to a choice between your life and the aggressor’s, you are, according to the principle of initiating behaviour, permitted to choose your own, and it would be wrong for everybody else to suggest that the deranged but innocent aggressor should be saved.24

It could be objected that the innocent aggressor and his or her victim are both innocent victims thrown together by circumstances for which neither is morally responsible, and that even if the innocent aggressor loses the immunity to attack, he or she retains a right to self-defence.25 A common objection to any reasonable belief view on justification is that it is likely to allow two parties to be justified in killing each other. But this objection does not apply to the idea put forward here. Two parties cannot both respond to each other’s initially threatening behaviour. It is part of the idea that one party takes the initial step, while the other responds to it.26 According to the principle of initiating behaviour, the innocent aggressor and his or her victim have not been thrown together. One party is thrown against the other and it is wrong of the former to thrust his or her bad luck upon the victim.

In this paper I have indicated that in order for an act of self-defence to be justified against an apparent aggressor or threat, the other person must initially have behaved in such a way—some physical movement, some verbal exclamation—that it is reasonable to interpret it as threatening. Given this

24. I said that the implications of being informed that the aggressor was innocence would be limited, not that it would be completely negligible. If you know that the aggressor is innocent you may have to endure more harm in order to avoid killing the aggressor than if he or she is culpable.
26. If they were pitted against each other, neither would be justified in killing the other. And it would be permissible to save either of the two; a procedure using lottery could be preferable.
account of justification, we can maintain the view that there is a significant relation between an act being justified and which party ought eventually to be given priority in a conflict situation. If we know that an act is justified, we know that we have to give preference to its agent when there is a choice between imposing equal harm on him or her and his or her victim. We should always give preference to the one who does not initiate the escalating process.

27. Another view on the matter could be to say that as long as the situation has not escalated, use of force is only excused. If the conflict escalates, the ultimately justified party is the one that did not initiate the situation.