

ARTICLE

---

# Why We Ought to be (Reasonable) Subjectivists about Justification

---

ANDREW BOTTERELL

The most difficult problems in criminal theory are generated by dissonance between reality and belief, between the objective facts and the actor's subjective impression of the facts.<sup>1</sup>

## I Introduction

Suppose Alice hits Bert on the head. *Prima facie*, Alice has done something that she ought not to have done: in hitting Bert she has harmed him, or violated a right of his not to be hit on the head, or interfered with his ability to determine his own ends.<sup>2</sup> And yet it may be that Alice should not be punished for hitting Bert, either because her action is justified, or because she is excused.

It has become commonplace to distinguish excuses from justifications and to point to an asymmetry between the two. According to one prominent view,<sup>3</sup> an excuse is something that calls attention to *features of the agent* at the time at which she performs an action. Somebody claiming to be excused does not deny that the action in question was wrongful, but denies that she was appropriately responsible for its performance. A justification, on the other hand, calls attention to *features of the situation or circumstances* at the time at which an agent performs an action. Somebody claiming a justification does not deny responsibility for the action in question, but denies that the action was wrongful in the circumstances.<sup>4</sup> More to the point, a justification entails that the accused has done nothing wrong, whereas an excuse entails that the accused has done something wrong, but that for various reasons her punishment should be reduced; or as Peter Westen puts it, "[t]he difference between justification and excuse, properly understood,

is as basic and simple as the distinction between, "I did nothing wrong," and, "Even if I did, it was not my fault."<sup>5</sup> This view of the distinction between justification and excuse is, broadly speaking, normative in nature, focusing on concepts of fault and responsibility.

According to another view, the difference between justification and excuse is related instead to the different role played by each in the criminal law. So, for example, Meir Dan-Cohen<sup>6</sup> distinguishes *conduct rules* (legal rules addressed to the public) from *decision rules* (legal rules addressed to officials) and suggests that justifications are conduct rules and are part of the theory of crime, whereas excuses are decision rules and are part of the theory of punishment. On this view, justifications have to do with whether a crime has been committed at all, whereas excuses are concerned with whether and to what extent the state may punish wrongdoers after the fact. In contrast to the normative analysis sketched above, this analysis of the distinction between justification and excuse is, broadly speaking, functional in nature.

I take no stand on which analysis is to be preferred, although my suspicion is that they are, at bottom, not so very different: justifications are conduct rules, and belong to the theory of crime, precisely because somebody who acts with justification does nothing wrong. Excuses are decision rules, and belong to the theory of punishment, precisely because somebody who asks to be excused has done something wrong but seeks exculpation on the grounds that she was not at fault for doing so.

Nonetheless, although the distinction between justification and excuse is reasonable enough as a point of

---

From July, 2007, Andrew Botterell will be an Assistant Professor in the Faculty of Law and Department of Philosophy, University of Western Ontario.

departure it leaves important questions unanswered, largely because it is not clear what facts are relevant to determining when an individual should be regarded as having acted with justification and when an individual should merely be excused for something that she did. It is not clear, in other words, whether an account of justification ought to be *subjective* or *objective* in nature. In "Competing Theories of Justification: Deeds v. Reasons" and elsewhere, Paul Robinson argues that justification ought to be understood in an objective manner.<sup>7</sup> In Robinson's view, the objective account of justification "generates liability results that are more just and that better match our collective intuitions of what is just," "lays bare the distinctions that are relevant to determining liability," and "allows a clearer analysis and a better perspective from which meaningfully to debate the competing issues."<sup>8</sup> If Robinson is correct, this has significant implications for how we ought to understand the nature of criminal liability. Still, despite the many compelling arguments marshaled by Robinson in support of his view, I have reservations about its overall coherence.

My aim is therefore to argue that justification should not be conceived of in purely objective terms. In arguing for that conclusion I will focus in particular on Robinson's presentation of that position, since it is the most sophisticated defense of the objective account of justification in the literature. My main point will be that the distinction drawn by Robinson between objective and subjective accounts of justification is problematic, and that careful attention to the role played by reasonable-

ness in subjectivist accounts of justification reveals that the apparent puzzles Robinson raises for subjectivism are merely apparent. I will suggest that we ought to be *reasonable* subjectivists about justification, where "reasonableness" is understood in a particular manner. This has consequences for various other issues, including how we make sense of mistaken justification, and I will have something to say about those issues as they arise along the way.

Although the issues I will discuss are not new, it seems to me that the objective view, and Robinson's particular presentation and defense of it, merits continued attention for several reasons. First, the objective view, if correct, has important ramifications for how we conceive of liability in criminal law in general and the nature of justification in particular. It is thus worth thinking about if we wish to understand the nature and function of criminal law. But second, independent of its role as an organizing principle in thinking about the nature of defenses, questions about the nature of justification and excuse are intrinsically interesting and theoretically challenging, and are therefore worth thinking about in their own right.

One final point: although I will be concerned throughout with justification and excuse, my examples will focus primarily on self-defense. Thus, my examples will focus on individuals defending themselves against the aggression of others; and I will be concerned with whether, and if so why, self-defense is justified in the circumstances in question.<sup>10</sup>

---

## II Why the Distinction between Justification and Excuse Matters

In my view the distinction between justification and excuse is an important one, and that remains the case regardless of whether we conceive of the distinction normatively, or whether we prefer to think about it in broadly functional terms instead. But the idea that the distinction is important has been challenged, with some claiming that it is idle, a distinction without a difference:

systems of classification into justifications and excuses suffer from a number of drawbacks. First, there is no agreement on the precise model that the classifications should take. . . . Secondly, there is no consensus as to which classification applies to which defense—for example, many see duress as excusatory but some as justificatory. Thirdly, there seems to be little agreement as to what difference, if any, would result in practical terms from classification of a particular defence into one category or another.<sup>10</sup>

Before continuing, therefore, let me try to address these concerns.

The issue of practical difference is best approached by looking at the situation of third parties and victims. For example, many believe that it makes a difference to the liability of a third party who is resisting an aggressor whether the aggressor is justified or merely excused. If the aggressor is justified, then resistance cannot be justified, whereas if the aggressor is merely excused resistance is permitted.<sup>11</sup> Similarly, there may be the possibility of conviction for aiding and abetting somebody who is acquitted by way of excuse, a possibility that is not available for an acquittal that proceeds by way of justification.<sup>12</sup> Finally, and relatedly, it pays to consider the situation of victims of justified or excused behavior. As I noted, it has often been said that a victim of justi-

fied action has no right to resist it, whereas the victim of an excused action does have such a right. But this has the further consequence that arguments for compensation are likely to be much more compelling in the case of victims of excused actions than they might be in the case of victims of justified actions.<sup>13</sup> Consequently, the distinction between justification and excuse can have important practical consequences for those acquitted of crimes, for those who aid or abet others, and for those who have themselves been the victims of crimes.

Moreover, even if the practical consequences are not always obvious, there remain theoretical reasons to suppose that the distinction between justification and excuse is important, even if the precise nature of the distinction is disputed, and even if there is disagreement about how, exactly, the classification applies.<sup>14</sup> I will confine myself to two points. First, I will suggest that the distinction is important if we are concerned about how we view ourselves, and about how others view us. And second, I will suggest that the importance of the distinction is not a function of its precision or ease of applicability.

First point first: consider the following remarks of John Gardner:

Criminal lawyers . . . tend to take it for granted that any doctrine that serves to acquit the accused, and therefore to avert the adverse normative consequence of her action, is as good as any other so far as the accused is concerned. I have always found this an astonishing assumption, which implies that nobody who is tried in the criminal courts has . . . any self-respect. . . . The self-respecting person aspires to live up to the proper standards for success in and fitness for the life she leads, and holds herself out to be judged by those standards. . . . She wants it to be the case that her actions were not truly wrongful, or if they were wrongful, that they were at any rate justified, or if they were not justified, that they were at any rate excused.<sup>15</sup>

Gardner's point is that there is a difference between saying that an individual did nothing wrong, and so should not be punished, and saying that an individual should not be punished, although she did something wrong. We want to be self-respecting individuals, and self-respecting individuals want their actions to be right actions, not merely actions that are excusable in retrospect. Similar remarks apply to how we would like others to regard and treat us. This suggests that the distinction between justification and excuse matters in a normative sense, since it matters to us what sort of individuals we wish to be and how we wish to be regarded and treated by others.

With respect to the second point, I will simply note that the demand for precision has critics of its own; indeed, some theorists have suggested that we should *not*

try to formulate a sharp distinction between justification and excuse precisely because such sharpness would distort the goals and purposes of the criminal law. Kent Greenawalt, for example, has argued that although the law's treatment of justification and excuse should generally track our understanding of comparable moral notions of responsibility and blame,

the law is necessarily crude. Its present imprecision is itself a caution that the law does not attempt to affix a precise moral label to each instance of behavior. Recognition of the goals the law can reasonably accomplish may be much healthier than pretentious aspirations to make the law the arbiter of every doubtful moral question.<sup>16</sup>

The fact is that moral language is more nuanced than legal language, as J.L. Austin noted when he referred to "terms, such as 'extenuation,' 'palliation,' 'mitigation,' [that hover] uneasily between partial justification and partial excuse."<sup>17</sup> Consequently, there will inevitably be moral distinctions having to do with culpability and responsibility that the law cannot adequately capture. And even if we are in agreement that it would be desirable if the law could, without cost, work sharp and clear distinctions between justification and excuse, it is not obvious that such rigidity and clarity is necessary in order for the distinction between justification and excuse to be generally comprehensible or useful.<sup>18</sup> And that, I suggest, is all that the law needs.

---

*The distinction between justification and excuse matters in a normative sense, since it matters to us what sort of individuals we wish to be and how we wish to be regarded and treated by others.*

---

In the end, perhaps the best can be said is this: if one accepts that the law requires some account of what constitutes a defense and some account of how various defenses differ in structure and effect, then one must agree that some organizing principle is wanted; and in my view, a principle that relies on a distinction between justification and excuse is as good a place to start as any. Consequently, I will in what follows assume that the distinction between justification and excuse is a distinction *with* a difference and that it matters both practically and normatively whether an actor is justified or merely excused.

### III Deeds and Reasons

As I said above, it is possible to think about justification in two contrasting ways, either subjectively or objectively. Robinson, however, prefers to use the terminology of *reasons* and *deeds* to capture this contrast.<sup>19</sup> Roughly put, the “reasons” theory of justification holds that a person *P* is justified in performing an action *A* in circumstances *C* if, given the world as *P* takes it to be, *P* believes that *A* is required in *C* in order for *P* to defend herself against aggression, to make an arrest, to preserve life, and so on. The key idea here is that justification is grounded in *subjective belief*: if *P* takes the world to be a certain way and if *P* believes that action *A* is required in the circumstances as *P* takes them to be, then *P* is justified in performing *A*.

For simplicity, let us say that if *P* believes that *A* is required in *C* to defend herself against aggression, then *P* has a *justificatory belief* that it is permissible for her to do *A* in *C*. Robinson’s interpretation of the reasons theory, then, can be rephrased as the claim that *P* is justified in doing *A* in *C* if *P* has a justificatory belief that *A* is necessary in *C*. Of course, much more needs to be said about what makes something a justificatory belief. Clearly, not just any subjective belief will do, since otherwise the notion of justification loses all of its normative force (I can have subjective beliefs for many reasons, some of them good, some of them bad). For the time being, then, I will use the phrase “justificatory belief” as a placeholder; I will return to discussion of what makes a belief genuinely justificatory below.

The “deeds” theory of justification, on the other hand, holds that what is important is not the world as *P* takes it to be, but rather the world as it actually is. Thus, on the deeds theory of justification a person *P* is justified in performing an action *A* in circumstances *C* if, given the world as it actually is, *A* is required in *C* in order for *P* to defend herself against aggression, to make an arrest, to preserve life, and so on. The key idea here is the *irrelevance* of *P*’s subjective beliefs to what she may or may not justifiably do. What is important is how things objectively are.

Now, in most cases of self-defense the deeds theory and the reasons theory generate the same liability results, which is to say that in most cases both theories agree on whether the individual acting in self-defense should be regarded as having acted with justification. This is because in most cases of self-defense, the way the world actually is *is* the way the agent in question takes the world to be. An example might be useful here.<sup>20</sup>

**Genuine Attacker:** Alice is in the park when she becomes aware of a club-wielding stranger, Bert, running towards her. Alice believes that Bert is about to attack her; and in fact, Bert is about to attack her.

In Genuine Attacker Alice is clearly justified in defending herself against Bert on both the reasons theory and the deeds theory. She is justified on the reasons theory because, given the world as Alice takes it to be—a world in which Bert is running towards her with a club, intending to cause her harm—her self-defensive actions are required in order for her to preserve her life. She has, in other words, a *justificatory belief* that defending herself against Bert is necessary in the circumstances. And she is justified on the deeds theory because, given the world as it actually is—a world in which Bert *is* running towards her with a club, intending to cause her harm—her self-defensive actions are also required in order for her to preserve her life. So far, so good.

Things become less clear, however, when appearance and reality diverge. Consider by way of illustration the following case:

**Mistaken Jogger:** Alice believes that Bert is running towards her with a club, and is intent on attacking her. In fact, Bert is simply out for a jog, and is holding a flashlight in his hand. Here Alice mistakes a jogger for an attacker.

As Robinson explains it, according to the reasons theory “a person will get a justification defence as long as he or she believes that the justifying circumstances exist. Whether they exist or not is irrelevant.”<sup>21</sup> Thus, on the reasons theory, Alice would appear to be justified in defending herself in Mistaken Jogger because she believes that the justifying circumstances exist: she believes, in other words, that Bert is about to attack her. And what if Alice acts as in Genuine Attacker but lacks such justificatory beliefs? On Robinson’s interpretation of the reasons view, it follows that Alice lacks a justification for her self-defensive actions. The explanation is simple: since, on the reasons view, justification is based on justificatory beliefs, it follows that in the absence of such justificatory beliefs there can be no justification.

The deeds theory, on the other hand, predicts that Alice lacks a justification in Mistaken Jogger. Says Robinson, “[u]nder the ‘deeds’ theory, whether the deed is in fact objectively justified is what matters; the actor’s reasons for acting are irrelevant to the justification defence (although they may be relevant to other doctrines of inculcation or exculpation.)”<sup>22</sup> The deeds theory maintains

that in order for Alice to have acted with justification, Alice must in fact *be* the subject of an attack. So according to the deeds theory Alice's self-defensive actions in

Mistaken Jogger are unjustified (which does not rule out the possibility that she may be excused, on which more later).

#### IV Justification and Harm

As Robinson presents it, the reasons theory is based on two core ideas. First, that an actor is to be regarded as having acted with justification if she believes that justificatory circumstances obtain, and second, that an actor is to be regarded as having acted with justification only if she acts on the basis of a justificatory belief. Robinson's critique of the reasons theory depends on this characterization of its structure. But as I hope to show, this characterization is in need of refinement. Before turning to those refinements, however, it is necessary to talk briefly about the relationship between harm and justification, since much of what I will say below depends on a particular view of how the law conceives of this relationship.

This brings me to a more fundamental idea at work in Robinson's defense of the deeds theory, namely that justification ought to be thought of in terms of net overall avoidance of harm.<sup>23</sup> According to Robinson, "[t]he test for justification . . . ought to be whether on balance the conduct in fact avoided a net societal harm (in the broadest sense of harm.)"<sup>24</sup> By way of illustration, consider the case of Motti Ashkenazi. Ashkenazi was an Israeli "who, with dishonest intent, took a back-pack he saw lying on a crowded beach. Unknown to Ashkenazi, the bag contained a terrorist bomb; inadvertently, he had saved many lives."<sup>25</sup> Robinson would presumably deny that Ashkenazi committed theft or did anything wrong in taking the backpack on the grounds that Ashkenazi's actions had a positively beneficial effect: "[i]f the justified act benefits society, it should be performed no matter what the motive and, further, if no harm has occurred, then the act is not a proper concern of the law."<sup>26</sup> This is because, on Robinson's view, in order to be regarded as having acted with justification an individual need not be aware that his actions will achieve a beneficial outcome so long as his actions *do* achieve a beneficial outcome, nor is it required that an individual have acted on the basis of a desire to do good; to the contrary, as with Ashkenazi, his motives may be quite bad indeed.

But why should we accept this view of the relationship between justification and harm? At one point Robinson says, in commenting on an argument of Kent Greenawalt, that

[a]fter insisting on viewing justifications under a "reasons" theory, in which mistaken justifications are deemed "justifications", [Greenawalt] reviews these "justifications", compares them to excuses, and concludes that the justification-excuse distinction is problematic. His premise, of course, assures his conclusion. The cat owner, who begins with the premise that his beloved Siamese really is his child in every meaningful sense, can then complain that the pet-child distinction is not nearly so clear as people think.<sup>27</sup>

Robinson is here concerned with Greenawalt's views about justification and excuse, and argues that Greenawalt's insistence on viewing justifications through the lens of the reasons theory is what leads him to question the boundary between justification and excuse. It seems to me, however, that precisely the same sort of complaint can be leveled against Robinson and his characterization and criticism of the reasons theory. Robinson begins with the idea that the point of criminal law is to minimize or avoid net overall harm. For simplicity, call this the "harm thesis." Based on this idea, he quite plausibly suggests that a person should be regarded as having acted with justification in the performance of some action *A* just in case *A* contributes to an overall reduction in harm. He then notes that on the reasons theory of justification, an action could be justified even if net overall harm is not minimized or avoided. This is because, on the reasons theory, Alice could be justified in harming Bert, say, given her belief that Bert is about to attack, and that could be true even if Bert is not about to attack her. This would result in a net overall harm, because Bert would suffer a harm despite the fact that he posed no threat to Alice, and no benefit would be realized, because as an objective matter Alice was not subject to attack in the first place. Robinson therefore concludes that the reasons theory must be mistaken since it misunderstands the nature of and the role played by the concept of justification in the criminal law.

More formally, Robinson seems to argue as follows:

- (P1) A person *P* is justified in performing an action *A* if doing *A* minimizes or avoids net overall harm.
- (P2) A person *P* cannot tell on the basis of her (subjective) beliefs whether an action *A* minimizes or avoids net overall harm.

(C) Whether an action *A* is justified cannot be a matter of what a person *P* believes.

And since the options, according to Robinson, are either a subjective reasons theory of justification or an objective deeds theory of justification, Robinson takes the foregoing to support the deeds theory.

My main point in what follows will be that the distinction between subjective and objective theories of justification is problematic. But let us set that aside for the moment and focus on the argument outlined above. I am prepared to allow that the inference from (P1) and (P2) to (C) is valid, and I am also prepared to grant (P2) since it seems very likely true. However, because I think that (C) is false, that gives me reason to think that there must be something wrong with (P1). So I would argue as follows: Robinson, after insisting on viewing criminal law in general, and justifications in particular, under the harm thesis, reviews the reasons theory's account of justification, and based on the fact that it is incompatible with the harm thesis, concludes that the reasons theory is problematic. His premise, of course, assures his conclusion.

The problem, as I see it, has to do with Robinson's focus on harm. I do not mean to deny that harm, in some form or another, has a role to play in the criminal law. But I doubt that it is the appropriate place to begin. For one thing, it is not clear what the harm thesis actually amounts to. What constitutes harm? How serious must a harm be in order for the criminal law to take an interest in it?<sup>28</sup> For another thing, it seems clear that the criminal law has a somewhat schizophrenic attitude toward harm: it permits some harmful actions even as it criminalizes other actions that are not harmful in any obvious way. By way of illustration, suppose Alice is suffering from great pain due to a terminal illness, but nonetheless refuses treatment. And suppose Bert, against Alice's wishes and motivated by a desire to help her, injects her with a drug that alleviates her pain. Here Bert has positively benefited Alice, and yet he is nonetheless guilty of assault.<sup>29</sup> Contrast this with a case where Alice is paralyzed by Bert in a sporting contest in which she was a

willing participant. Here Bert has surely harmed Alice, and yet he is innocent of wrongdoing.<sup>30</sup> Or again:

If I deprive you of something to which you have no right—the view of the sunset over my land, or the use of someone else's property—I do not wrong you. If I deprive you of something to which you have a right, I wrong you even if I do not harm you. Suppose that, without consulting you, I take an afternoon nap in your bed, while you are out of the house. (I bring my own sheets, sweep the floor on my way out, and so on.) I do you no harm, but I wrong you in Kant's sense.<sup>31</sup>

The point is that the law does not protect us from injuries or harms that arise out of our being deprived of something to which we have no right, even as it seeks to protect us from harmless wrongs.<sup>32</sup> This suggests that there is something problematic with the idea that harm is the lynchpin upon which criminal law pivots.<sup>33</sup>

It should come as no surprise, then, that the deeds theory—a theory of justification that takes as its starting point the notion of net overall harm or benefit—leads to liability results that are incompatible with the reasons theory—a theory of justification that is based on other considerations. If what is at issue in questions of justification is the minimizing or avoidance of net overall harm, where harm is understood objectively, then it is not implausible to ask whether cases of putative self-defense are genuine cases of justification, whether they do, objectively speaking, minimize net overall harm. But if our starting point is not harm, but is instead something like autonomy, or freedom of choice, or equality of and equal respect for the rights and interests of others, then an account of justification that is primarily concerned with whether individuals are treated equally will surely be more plausible. That a reasons theory based on such concepts and principles conflicts with an account of justification based on the minimization or avoidance of harm is not surprising. But neither does it show that the reasons theory is thereby mistaken. It simply highlights the fact that reasons theorists about justification should not be harm theorists about the criminal law. My argument in what follows is best viewed as an extended comment on this somewhat simple idea.

---

## V The Structure of the Reasons Theory

I will be arguing that we ought to prefer a suitably supplemented reasons theory to Robinson's deeds theory. Still, it should be granted at the outset that the idea that something other than subjective belief might determine

the conditions for criminal liability is not implausible. Take the following example from George Fletcher. Consider, says Fletcher,

a case in which Williams marries a second time thinking that he is still married to his first wife. In fact, his estranged wife has obtained a valid, *ex parte* divorce in another jurisdiction. Though Williams is legally single, he thinks his second marriage is bigamous. Is he in fact guilty of bigamy? In our legal tradition . . . he is patently not guilty. We may take it as common ground that the prior valid divorce, whether Williams knows about it or not, is sufficient to bar conviction for bigamy.<sup>34</sup>

Regardless of what Williams might subjectively believe, the objective fact that he is divorced is sufficient to render his actions non-bigamous. Similarly, consider the case of Ora H. White, who was charged with abandoning his wife while she was pregnant.<sup>35</sup> While admitting that he had abandoned her, White argued that he should not be found guilty because he was unaware that she was pregnant. A jury convicted White on the grounds that his subjective belief as to his wife's condition was irrelevant to his culpability, and that verdict was upheld on appeal. So it would appear that there are reasons to think that objective conditions might sometimes be more important than subjective beliefs in determining what is legally permissible. The question is whether objective conditions are more important than subjective beliefs when justification is at issue. This is what the reasons theory denies.

What, according to the reasons theory, is the connection between reasons and justification? According to Robinson,

[t]he standard formulation [of the reasons theory] provides that "an actor is *justified* if he *believes* that the conduct is necessary to" defend against unlawful aggression, to make an arrest, to maintain order on the vehicle, and so on. Under the "reasons" theory, a person will get a justification defence as long as he or she believes that the justifying circumstances exist. Whether they actually exist or not is irrelevant.<sup>36</sup>

This suggests that the reasons theory is committed to a certain sufficiency claim: believing that there are justificatory reasons is sufficient for having a justification. In short:

**Sufficient Reasons Theory:** *P* is justified in doing *A* in circumstances *C* if *P* has a justificatory belief that *A* is required to avoid harm to self in *C*.

But Robinson also says that "if the justifying circumstances do exist but the actor is unaware of them and acts for a different purpose, the 'reasons' theory denies a justification defense."<sup>37</sup> In short: if the actor lacks a justificatory belief, then the actor lacks a justification. And so by contraposition: if the actor acts with justification, then the actor must have a justificatory belief. This passage therefore suggests another view about the relation between justificatory beliefs and actions:

**Necessary Reasons Theory:** *P* is justified in doing *A* in circumstances *C* only if *P* has a justificatory belief that *A* is required in order to avoid harm to self in *C*.

The upshot is that the view that Robinson is attributing to the reasons theorist must be that having a justificatory belief is both necessary and sufficient for having a justification. So Robinson must be attributing the following view to the reasons theorist:

**Necessary and Sufficient Reasons Theory:** *P* justified in doing *A* in *C* if and only if *P* has a justificatory belief that *A* is required in order to avoid harm to self in *C*.

Now, it may be that many reasons theorists would be happy enough with this view. But what I want to point out is that it is not mandatory: somebody could be a reasons theorist while rejecting the Necessary and Sufficient Reasons Theory and endorsing the Necessary Reasons Theory alone. For example, somebody could say that having a justificatory belief is necessary for having a justification, but remain agnostic about what else might be required in order to make a belief genuinely justificatory. One might wonder whether this should qualify as a reasons theory at all, since according to it reasons are not sufficient for justifications, and since it would seem to preclude a reductive analysis of justification in terms of beliefs or reasons. But this objection is unpersuasive, since it isn't clear that a necessary reasons theorist must be concerned to offer a reductive analysis of "justification" at all. Still, because I suspect that most reasons theorists would be prepared to adopt something like the Necessary and Sufficient Reasons Theory, it is on that view that I will focus my attention.

## VI The Reasonable Belief Requirement

According to Robinson, the reasons theory is based in part on the idea that an actor is to be regarded as acting with justification if she believes that justificatory circumstances obtain. But this needs refinement. This is

because it neglects an important issue, namely the *kind* of belief that is relevant to an agent's having acted with justification. Robinson talks about an actor's belief that conduct is necessary to defend against aggression. But

what Robinson ought to talk about instead is an actor's *reasonable* belief that conduct is necessary to defend against aggression.

To see how the introduction of a reasonableness component might be relevant to the sorts of issues raised by Genuine Attacker and Mistaken Jogger, imagine the following hypothetical scenario:

**Indecisive Attacker:** Alice believes that Bert, who is running towards her with a club, is intent on attacking her. Bert, however, is indecisive and hasn't decided what he is going to do: he may run past Alice, or he may attack her. To keep his options open, however, Bert continues to run towards Alice, club in hand.

By hypothesis Bert is not presently attacking Alice. To be sure, he *may* attack Alice; but then again, he may not. It follows that on the deeds theory, Alice would act without justification were she to use defensive force to protect herself. Since she is not presently being attacked and since she may not actually be attacked at all, the requisite justifying circumstances fail to obtain, and thus the deeds theory says that her self-defensive actions are unjustified.<sup>38</sup> It seems to me, however, that Indecisive Attacker is a case where self-defense is clearly permissible. And this is because Alice's belief that she is being attacked in Indecisive Attacker is a *reasonable* one in the circumstances.

The problem that Indecisive Attacker presents for the deeds theory is this: Alice acts reasonably in defending herself against Bert. Yet because on the deeds theory the requisite justifying circumstances fail to obtain, it turns out that Alice's self-defensive actions, while reasonable, are unjustified. This means that in defending herself in Indecisive Attacker Alice acts badly, and that remains the case even if she is later excused. But if Alice acts permissibly in Genuine Attacker, how can her equally reasonable response in Indecisive Attacker fail to justify her self-defensive actions in that case? The problem is that the deeds theory seems to get Indecisive Attacker wrong.

In response it might be pointed out that this objection assumes that actions can be justified only if the justifying circumstances presently obtain, and it is arguable that this is not something that a deeds theorist must accept. For clearly, one of the points of the deeds theory is that what makes an action justified might obtain only after the action has been performed, when its salutary benefits have been weighed against its deleterious effects. A better example might therefore be the following:

**Impossible Attacker:** Bert is running towards Alice with a club, intent on attacking her. Unbeknown to either Alice or

Bert, Bert has badly clogged arteries. If Bert continues to run towards Alice, the continued exertion will result in Bert's suffering a stroke just before he reaches her, and his threat will be neutralized.

In Indecisive Attacker it is compatible with Bert's running towards Alice that he might attack her, whereas in Impossible Attacker it is *incompatible* with Bert's running towards Alice that he might attack her. This means that the objective justifying circumstances will never obtain. Yet it surely cannot be incumbent on Alice to wait and see whether Bert is going to attack her before she is permitted to act in self-defense. As Justice Holmes famously put it, "detached reflection cannot be demanded in the presence of an uplifted knife."<sup>39</sup> But this seems to be precisely what the deeds theory requires of people. The deeds theory says that Alice acts without justification if she defends herself in Impossible Attacker—since no attack will happen, nothing good can come from Alice's self-defensive actions—and that is a strike against it. The point is not that in Impossible Attacker we believe that Alice should not be punished; that much is presumably common ground between both reasons and deeds theorists, although their explanations for the lack of punishment will differ, the reasons theorist saying that Alice acted with justification, the deeds theorist saying that Alice is excused instead. Rather, the point is that in Impossible Attacker we believe that Alice did not act badly *at all*. She does not need to avail herself of an excuse because she did nothing wrong; she does not need the benefit of a decision rule because her actions do not engage any conduct rule. Moreover, if a friend of the deeds theory responds by saying that before defending herself Alice should wait until just before the point at which her self-defensive actions would not save her, then we have abandoned the deeds theory altogether, since we are now no longer concerned with whether the objective justifying circumstances in fact exist.

Still, it might be objected that the addition of a requirement of reasonableness cannot make much of a difference: if the reasons theory runs into trouble when it is belief *simpliciter* that is at issue, then surely, it might be thought, it will run into trouble when it is *reasonable* belief that is at issue. For if the problem is that Alice's belief that she is being attacked by Bert is compatible with her not being attacked by Bert, then this will be a problem on the reasonable belief view as well, since Alice's *reasonable* belief that she is being attacked by Bert is also compatible with her not being attacked by Bert. To the contrary, however, I think that focusing on the reasonableness of Alice's belief bears directly on the issue between reasons and deeds theorists, as I will now argue.



## VII Reasonableness as Equality

I said earlier that I would be using the phrase “justificatory belief” as a placeholder. It should now be clear what place it was holding: a justificatory belief is a reasonable belief. But what is a reasonable belief? It is tempting to view a reasonable belief simply as an internally coherent belief, a belief that makes sense from the agent’s own subjective perspective. Relying on an account of reasonableness developed by Arthur Ripstein and others, however, I will suggest that we ought to interpret reasonableness in a different way.<sup>40</sup>

But before doing so, however, let me say something about why I think the general sort of account of reasonableness developed by Ripstein is worth taking seriously. It must be acknowledged that Ripstein’s account is not free from problems.<sup>41</sup> For one thing, any account that attempts to make sense of the way the concept of reasonableness functions in the law must be capable of being generalized in a plausible way, and some have questioned whether Ripstein’s account can be thus extended.<sup>42</sup> For another thing, although on Ripstein’s account it must be the case that somebody who acts reasonably, or on the basis of a reasonable belief, acts with justification, “reasonable” cannot simply mean “justified” on pain of making the reasons theory tautological, amounting to nothing more than the claim that somebody acts justifiably so long as she acts with justification. But because it has been argued that any attempt to explicate a notion of reasonableness using language other than the language of justification is bound to fail, this is another concern.<sup>43</sup> So why am I choosing to focus on the account of reasonableness suggested by Ripstein?

One reason is simple: Ripstein’s is a sophisticated account of the nature of reasonableness implicit in the law, and to the extent that some theory of reasonableness is required in order to make sense of the reasons theory, Ripstein’s would seem to be precisely the sort of account that is needed. But an even better reason, it seems to me, is the following. Robinson’s presentation of the distinction between reasons theories and deeds theories rests on a more fundamental distinction between the subjective and the objective. In its treatment of justification, for example, the deeds theory screens off as being irrelevant matters having to do with the agent’s own subjective beliefs. If, however, that more fundamental distinction turns out to be problematic, Robinson’s distinction between deeds and reasons theories of justification will itself be called into question. Because Ripstein’s account

of reasonableness nicely problematizes that very distinction, it ought for that reason to be considered closely. And finally, even if one is unhappy with the precise way in which the concept of reasonableness is developed by Ripstein, it remains the case that any account of reasonableness worth its name will distinguish merely subjective belief from genuinely reasonable belief, and so will address the same issues treated by Ripstein’s account. Consequently, there seems to me to be no harm in appealing to Ripstein’s account of reasonableness to underscore the general points I wish to make concerning the reasons theory.

With those preliminaries out of the way, let me turn to discussion of the concept of reasonableness in which I am interested. In Ripstein’s view, there is an important distinction to be drawn between *rationality* and *reasonableness*. Ripstein finds this distinction in the work of John Rawls.<sup>44</sup> According to Rawls, reasonable persons “desire for its own sake a social world in which they, as free and equal, can cooperate with others on terms all can accept.”<sup>45</sup> Merely rational persons, on the other hand, lack “the desire to engage in fair cooperation as such, and to do so on terms that other as equals might reasonably be expected to endorse.”<sup>46</sup> This does not mean that reasonable persons are not rational persons. To the contrary, being complementary ideas neither the rational nor the reasonable can stand without the other: “[m]erely reasonable agents would have no ends of their own they wanted to advance by fair cooperation; merely rational agents lack a sense of justice and fail to recognize the independent validity of the claims of others.”<sup>47</sup> It does mean, however, that for certain purposes reasonableness should be viewed as being conceptually prior to rationality, since it is against the backdrop of the concept of reasonableness that the imposition and assumption of benefits and burdens should be measured.

Building on Rawls’s distinction, then, we can say that a rational person is a person who acts to further her own ends, where such ends are often conceived of in economic terms (although they need not be). This sort of person behaves in an instrumental manner: she asks herself what she should do given the ends that she has. A reasonable person, on the other hand, acts in a way that first and foremost recognizes the freedom of others: she asks herself what she should do given the competing interests of others and given her desire to interact with others on terms all can accept. The idea of reason-

ableness is thus what might be called a co-operative idea. It is also a practical one, as opposed to an abstract capacity to reason with respect to theoretical issues. So when I talk about reasonableness I mean to talk about “the style of deliberation a person would ideally engage in, and the impartial attention he would give to competing values and evidences in the concrete setting.”<sup>48</sup>

It should be clear, moreover, that the concepts of rationality and reasonableness diverge. To see why, assume, with Gary Schwartz,

that there were no liability for negligence. The Rational Man would behave negligently, and economics would afford no reason to consider him morally blameworthy for doing so; the Reasonable Man, by contrast, would conduct himself in a nonnegligent fashion.<sup>49</sup>

The explanation is simple: because the Rational Man acts to further his own ends, in the absence of liability for negligence there would be no countervailing reason for him to act nonnegligently. The Reasonable Man, on the other hand, would continue to act nonnegligently simply because that is the way reasonable individuals behave; in the view of the Reasonable Man, negligent action is “subnormal and deviant.”<sup>50</sup>

Rawls’ co-operative idea of reasonableness is Kantian in nature, in the following sense. According to Kant, the universal law of right is to “act externally [so] that the free use of your choice can coexist with the freedom of everyone in accordance with a universal law[.]”<sup>51</sup> On Kant’s view, one acts permissibly—that is, within one’s rights—so long as the free exercise of one’s choice is compatible with the free exercise of the choices of others. If my acting in some way means that you are prohibited from freely acting in some other way, then my action does not respect, and so is incompatible with, your freedom. In short: I act reasonably if my actions can co-exist with your freedom. And my actions can co-exist with your freedom if I act in a way that treats your rights and interests as worthy of equal concern. So I act reasonably if I act in a way that treats your rights and interest as worthy of equal concern. This is, in a nutshell, the conception of reasonableness in which I am interested. For simplicity I will follow Ripstein in calling this the idea of *reasonableness as equality*.

A similar conception of reasonableness has been championed by Gregory Keating, who has proposed an account of *reasonableness as reciprocity*. Relying, like Ripstein, on the work of Kant and Rawls, Keating argues for an account of reasonableness according to which “[t]he risks that any given individual should subject herself to are properly determined by criteria of (individual) ra-

tionality; the risks that each of us should be entitled to impose on each other are properly determined by criteria of (interpersonal) *reasonableness*.”<sup>52</sup> Because, says Keating,

due care as reasonableness holds that the lives of persons are distinct and their aims and aspirations different . . . it argues that the law of reasonable risk imposition is best understood as an attempt to identify those precautions that serve as natural focal points for reasonable persons seeking to sustain a mutually beneficial form of social interaction.<sup>53</sup>

Keating focuses on the difference between rational and reasonable behavior, and suggests that reasonable persons are those who desire to “sustain a mutually beneficial form of social interaction.” Because such a form of social interaction will succeed only to the extent that the rights and interests of others are treated with equal concern, Keating’s conception of reasonableness as reciprocity shares the same conceptual architecture as the conception of reasonableness as equality.

---

*I act reasonably if I act in a way that treats  
your rights and interest as worthy  
of equal concern.*

---

In any event, the idea of reasonableness as equality (or reciprocity) leads naturally to issues concerning risks. This is because in a world of risks the idea of equality can be developed in a way that directly bears on the question of what constitutes a fair allocation and assumption of risks. It has long been noted that risks are part of life: as Lord Reid said in *Bolton v. Stone*, “[i]n the crowded conditions of modern life even the most careful person cannot avoid creating some risks and accepting others.”<sup>54</sup> But though we create and assume risks all the time, the important question is when the creation or acceptance of a risk is reasonable. In assessing what risks are reasonable to impose or accept, we are implicitly asking where, as among individuals, risks and their consequences should lie.

Tort law provides us with a fantastic array of risk allocation problems. A classic formulation of such problems can be found in Judge Learned Hand’s decision in *United States v. Carroll Towing Co.*<sup>55</sup> There the question was whether the owner of a barge owed a duty to keep an attendant on board while the barge was moored in the harbor. Judge Learned Hand assessed the owner’s

duty in terms of three variables:  $P$ , the probability that the barge might break away from its moorings;  $L$ , the gravity of the resulting injury or loss were the barge to break away; and  $B$ , the cost of precautions sufficient to prevent the barge from breaking away. According to Hand, the owner would be negligent if  $B < PL$ , if, that is, the burden of precautions was less than the probability of harm times the gravity of the harm, and if the owner failed to take precautions.

---

*Risk cannot be reasonably allocated where one party unilaterally and indifferently sets the terms of the interaction with another.*

---

The Hand formula provides a nice way to understand the contrast between reasonableness and rationality. I might decide to take only those safety precautions that maximize my expected utility by minimizing my expected accident-related costs; indeed, if the cost of precautions is greater than any expected accident-related costs I might forgo precautions altogether. However, although such an approach might be *rational* it might fail to be *reasonable* since it might not take into account the interests and freedom of others. As Ripstein puts it:

if I am able to inexpensively limit my precautions to those that protect people who are likely to sue, . . . I behave rationally but not reasonably. In contrast, if I take such precautions as are justified by an appropriate balance between the interests in liberty and security that all are supposed to share, I behave reasonably.<sup>56</sup>

It is difficult to say in the abstract what behavior is reasonable, and which risks ought reasonably to be imposed and accepted. The reasonableness of an imposed or accepted risk will depend on the liberty and security interests at stake, the consequences that would follow should things go wrong, and the extent to which those consequences are foreseeable. If the consequences for you of my driving my car quickly are remote and minimal, then it might not be unreasonable for me to do so. If the consequences are foreseeable and dire, on the other hand, then it would be unreasonable for me to prefer my interest in my own liberty to your interest in your own security. Such a preference would not be reasonable since it would not treat your rights and interests as worthy of equal concern.

Among the things that follow from the idea of reasonableness as equality is that risk cannot be reasonably allocated where one party unilaterally and indifferently sets the terms of the interaction with another.<sup>57</sup> The idea that individuals may not unilaterally and indifferently set the terms of their interactions with others makes problems for the deeds theory. For if the deeds theory of justification is adopted, a person is justified in defending herself only if an attack is *in fact* occurring. But this means that the aggressor is unilaterally and indifferently setting the terms of the interaction, and unilaterally and indifferently imposing risks on another person, because it is up to the aggressor to decide whether an attack is, after all, going to occur. With this in mind, it seems unfair to insist that the potential victim must wait until an attack is in fact occurring before she can take steps to defend herself. Of course, it might again be replied that since no one knows at the time self-defensive force is deployed whether the justificatory circumstances obtain, the deeds theory gives no answer *ex ante* to whether the victim's deployment of self-defensive force is justified. However, this understates things: the deeds theory does tell us something, since it suggests that the victim's right to be free from unconsented-to risk counts for less than the aggressor's right to impose such risks on others. And that seems questionable. Again, this is because from the standpoint of equality one cannot prefer one's freedom of action to another's interest in security unless one has reasonably taken that security interest into account. Failure to do so means that one is liable for the consequences of one's actions. Reasonable risks and their attendant consequences lie with others; unreasonable risks, together with any resulting injuries, belong to those who create them. And this is what the deeds theory fails to acknowledge. Because Bert acts unreasonably in Indecisive Attacker, the risk that Alice might deploy self-defensive force and injure Bert belongs to him, and not to Alice. As Kant might say, in acting unreasonably and unilaterally and indifferently imposing a risk on Alice, Bert plays a game of chance with Alice's agency.<sup>58</sup>

Conversely, were the victim's mere subjective belief that she is being attacked sufficient to justify her self-defensive actions, then *she* would be able unilaterally to set the terms of her interaction with her attacker, because nothing the attacker could do would affect whether the victim had a justificatory belief. And this seems equally unfair, because nothing the would-be attacker can do has any bearing on the proper assessment of the victim's actions. I take these observations to show that we need something like a notion of reasonableness as equality in

order to navigate between these two extremes. For if we understand reasonableness in terms of equality, then we can say that Alice is justified in defending herself against Bert if she reasonably believes that Bert is about to attack her, where reasonableness is not merely a matter of what Alice subjectively believes and is not simply determined by objective facts about the world, but is instead a matter of what a suitably placed observer would conclude about the circumstances in question.

Let me be clear: I am not suggesting that the only thing that makes a criminal assault wrongful is the fact that the attacker is unilaterally and indifferently setting the terms of his interaction with his victim. For one thing, such a view seems to inadequately capture the wrongfulness of assault.<sup>59</sup> For another, it seems to suggest that if only victim and attacker had acted bilaterally, no criminal wrong would have occurred, and it is not entirely clear what it might mean to act bilaterally in the context of the criminal law.<sup>60</sup> Finally, without a way to rank unilateral interactions so that some are treated as being worse than others, it seems to view all criminal wrongs as being of equal standing. Rather, my point is that there is an important connection between unreasonableness, wrongful action, and the right to act so as to protect one's own security: by acting unilaterally and indifferently and by failing to take the rights and inter-

ests of others seriously, one assumes a certain risk. This is why—in Indecisive and Impossible Attacker—Bert must bear the costs of Alice's reasonable responses to his unreasonable actions.

To recapitulate, I have been arguing that the distinction drawn by Robinson between purely subjective "reasons" theories and purely objective "deeds" theories is problematic. Reasons theories should be concerned with *reasonable* beliefs, where reasonableness is understood in terms of equality. Moreover, whether an actor pays due attention to the rights and interests of others—whether, that is, risks are properly allocated—is not a purely subjective matter; to the contrary, it necessarily involves external, public considerations. As a result, the reasons theory is most charitably reinterpreted as follows: *P* is justified in doing *A* in *C* if and only if *P* has a *reasonable* belief that *A* is required to avoid harm to self in *C*. In my view, this sort of reasons theory bridges the gap between subjective and objective theories of justification. It remains subjective insofar as it insists that somebody claiming a justification must have certain beliefs about the way the world is, but it goes beyond mere subjectivity in insisting that the belief be a reasonable one, where reasonableness is conceived of from the standpoint of equality.

---

## VIII Justification and Reasons for Acting

Thus far I have been arguing that Alice's belief that Bert is threatening her must be a reasonable one in order for self-defense to be justified. But this raises a further issue concerning *how* Alice's self-defensive actions must be related to her belief that Bert is a threat. I said earlier that, as Robinson presents it, the reasons theory is based on two core ideas. First, that an actor is to be regarded as having acted with justification if she believes that justificatory circumstances obtain. And second, that an actor is to be regarded as having acted with justification only if she acts on the basis of a justificatory belief. I suggested some reasons for viewing the first idea with skepticism. I now want to discuss the second idea. Thus, we need to ask whether Alice must be acting on the basis of her reasonable belief that Bert is threatening her in order for her self-defensive actions to be justified, and if so, what that claim amounts to.

As we have seen, sometimes Robinson suggests that underlying the reasons theory is the claim that if an agent has a justificatory belief, then that agent has a jus-

tification. Call this the *justification as belief* interpretation of the reasons theory. But at other times he seems to suggest that underlying the reasons theory is the stronger claim that in order for an agent to be justified, she must not only have a justificatory belief, but further, that she must act on the basis of that justificatory belief. Call this the *reasons for action* interpretation of the reasons theory. As he says, "[i]f the 'reasons' theory of justification truly lay at the root of present justification defenses, those defenses would require more than a belief in the justifying circumstances. They would require that the actor's purpose was the justificatory one."<sup>61</sup>

The following hypothetical nicely illustrates the distinction. Suppose Alphonse wants to hurt Buford (his reasons for doing so do not matter), but has not done so for fear of punishment. One day, however, Alphonse sees Buford attacking an innocent bystander. Alphonse intervenes, beating Buford. By hypothesis, Alphonse is "motivated not by a desire to protect the victim but rather by his desire to hurt Buford without risking li-

ability."<sup>62</sup> Robinson asks,

[d]oes Alphonse deserve a justification under the rationale of the "reasons" theory? Is his conduct "morally proper"? Are his 'reasons' for pummeling Buford "sound and good"? No. His reasons for acting are base indeed: his long simmering hatred. Yet, he nonetheless will get a defence under the typical "believes" formulation of current law.<sup>63</sup>

This example gives rise to a puzzle, to be sure, but it also seems to conflate two distinct issues. This is because it runs together our moral assessment of Alphonse and his actions with the law's assessment of what is permissible. The law is not primarily concerned with whether Alphonse's conduct is "morally proper." Instead, it is concerned with whether, given the world as Alphonse takes it to be, Alphonse should be regarded as having acted permissibly. In short, it is not clear what *legal* consequence ought to follow from our negative *moral* assessment of Alphonse and his actions.

But let us set this worry to one side. Robinson's point seems to be that a reasons theorist must endorse the idea that in order for Alphonse to be justified in beating Buford, his reason for acting must be appropriately related to his justificatory belief that Buford is unlawfully attacking an innocent bystander. In other words, Robinson's first point seems to be that reasons theorists must accept the reasons for action interpretation of the reasons theory. But his second point is that this is in tension with the typical "believes" formulation of the reasons theory. For the "believes" formulation predicts that Alphonse should be regarded as having acted with justification, whereas the reasons for action interpretation predicts that he should not be so regarded. The upshot, according to Robinson, is that there is something inconsistent about the reasons theory. But must a reasons theorist accept the reasons for action interpretation of the reasons theory? It seems to me that a reasons theorist has three options. First, she can deny that reasons theorists must accept the *reasons for action* interpretation of the reasons theory and argue that they can stick with the *justification as belief* interpretation instead. Second, she can accept Robinson's first point, but argue that Alphonse's reason for beating Buford *is* appropriately related to his justificatory belief. Or third, a reasons theorist can accept Robinson's point about the connection between reasons and actions, but insist on a more robust account of what makes something an appropriate reason for action and deny that Alphonse is justified in beating Buford.

Now, I think that the first option—rejecting the *reasons for action* interpretation of the reasons theory while retaining the *belief as justification* interpretation—is, though

not incoherent, at least ill advised. This is because Robinson is surely correct in supposing that a plausible reasons theory will link justificatory beliefs to reasons for action. In particular, a plausible reasons theory will presumably hold that an individual will be regarded as having acted with justification only if her reasonable belief that her actions were justified was somehow implicated in her acting in the manner in which she did. In other words, it seems to me that a reasonable reasons theorist should adopt the reasons for action interpretation of the reasons theory. So I propose to set this first response aside.

---

*Robinson is surely correct in supposing that a plausible reasons theory will link justificatory beliefs to reasons for action.*

---

What of the second response? A reasons theorist must surely agree that in order for his beating of Buford to be justified Alphonse must reasonably believe that Buford is unlawfully attacking the victim. But a reasons theorist need not agree that Alphonse's actions will be justified only if his reasonable belief that Buford is attacking an innocent stranger was the sole reason for him so acting. According to this response a looser connection between beliefs and reasons for action suffices. A rough "but for" test is helpful here. Ask: would Alphonse have attacked Buford if he had not had a justificatory belief that Buford was attacking an innocent bystander? If the answer is yes, then Alphonse's justificatory belief is explanatorily idle, and so is not appropriately related to his actions, in which case his actions are not justified. If the answer is no, then there is reason to think that Alphonse's action is appropriately related to his justificatory belief and, hence, that his belief *is* causally relevant to his self-defensive actions. In other words, if Alphonse says, yes, I beat Buford because of my long-simmering hatred for him, but I would not have beat him had I not seen him attacking an innocent victim, then Alphonse is saying that his actions were counterfactually connected to his justificatory belief: had he not had that justificatory belief, he would not have beaten Buford. On this view, that is all that is required in order for Alphonse's actions to be appropriately linked to his reasonable justificatory belief.<sup>64</sup>

This will no doubt seem puzzling, since it appears to leave open the possibility that a person could be justi-

fied in acting in self-defense even though that person does not act *solely in order to defend herself*. But I would argue that if this is indeed puzzling it is puzzling mostly for linguistic reasons. To say that I acted in self-defense typically entails that I acted only in order to preserve my own life. But if the foregoing response to Robinson is taken seriously, then in some situations this entailment will fail, since in some situations beliefs and reasons for action will come apart. Once it is allowed that a weaker connection between justificatory beliefs and reasons for action is possible, however, then it should come as no surprise that justificatory beliefs might diverge from reasons for action in certain cases.<sup>65</sup> In any event, according to this response, because Alphonse's actions are appropriately linked to his justificatory belief that Buford is assaulting an innocent bystander, he should be regarded as acting with justification. On this view, in other words, the typical "believes" formulation of the reasons theory gets the example of Alphonse and Buford exactly right.

The third response is one that argues that only certain kinds of justificatory beliefs can function as justifications. It therefore predicts that in Robinson's hypothetical Alphonse is *not* justified in beating Buford because he does not act on the basis of the right kind of belief. This response is suggested by John Gardner. Gardner asks the following question: "[a]re one's actions and beliefs justified by the reasons which actually applied to one, or by the reasons which, perhaps mistakenly, one thought applied to one and accordingly treated, in one's acting or believing, as if they were reasons which actually applied to one?"<sup>66</sup> Gardner says that in order for an individual to be justified in acting in a certain manner, it must be true both that "there was an applicable (guiding) reason for so acting . . . and that this [reason] corresponded with the (explanatory) reason why the action was performed."<sup>67</sup> A guiding reason, according to Gardner, is a fact or reason that grounds what one ought to do or believe; an explanatory reason is a reason on the basis of which one in fact acts or comes to believe. So, for example, I might come to believe that there is an apple in front of me because it perceptually appears to me that there is an apple in front of me. And it might perceptually appear that there is an apple in front of me because there is a red, round sense-datum in some part of my visual field. That it perceptually appears that there is an apple in front of me would be an explanatory reason for my coming to believe that there is an apple in front of me. But as many philosophers have argued, it is an open question whether such a perceptual state could be a guiding reason, a reason on the basis of which I ought to acquire beliefs about the external world. Thus

on Gardner's view, an individual is justified in acting in a certain way just in case she acts on the basis of an explanatory reason that is appropriately linked to a guiding reason: the reason on the basis of which such an individual acts must be a reason on the basis of which one ought to have acted.

By way of illustration, consider Gardner's example of a soldier stationed at a checkpoint in a war-torn region who misinterprets some motorist's actions—the soldier thinks the motorist is reaching for a weapon when she was merely reaching for identification documents—and shoots and kills her.<sup>68</sup> This is a subjective belief about the way the world is: the soldier believes that the world is one in which the motorist is about to pull out a weapon. But the issue is whether that belief is reasonable in the circumstances. If it is, then it could constitute a guiding reason for his action. Why? Because his belief would not be merely subjective, but would instead reflect the fact that the motorist's actions were actions that, on any reasonable interpretation, unfairly imposed a risk on the soldier.

---

*A guiding reason, according to Gardner, is a fact or reason that grounds what one ought to do or believe; an explanatory reason is a reason on the basis of which one in fact acts or comes to believe.*

---

This analysis would apply to the case of Alphonse and Buford as follows: in order for Alphonse to be justified in beating Buford, (1) its reasonably appearing to Alphonse that Buford was threatening an innocent bystander must be a guiding reason for beating Buford; and (2) Alphonse must have acted on the basis of that reason. Since, in Robinson's example, Alphonse did not act solely on the basis of his reasonable belief, but acted as well on the basis of his hatred for Buford, Gardner would say that Alphonse acted without justification: his reason for acting was not a guiding reason. In short, this response accepts Robinson's point about the relation between justificatory beliefs and reasons for actions, but argues that the "typical believes formulation of justification" needs to be supplemented by something like Gardner's distinction between guiding and explanatory reasons. This response does not abandon the reasons theory, but concedes that it requires further development and supplementation.

Reasons theorists therefore have a choice. They can either adopt a weaker conception of what it means to act on the basis of a justificatory belief, or they can impose more stringent conditions on what makes a belief truly justificatory, and claim that in some situations the typical “believes” formulation gets the analysis of justification wrong. Which option should reasons theorists prefer? Although this is a difficult question, I am inclined to adopt weaker constraints on the notion of acting on the basis of a justificatory belief. My reason for doing so is that I have some reservations about Gardner’s proposal understood as a true supplement to the reasonable belief version of the reasons theory. This is because it seems to me that a guiding reason is nothing more than a reasonable belief, and so talk of guiding reasons

adds nothing to the analysis of justification proposed by the reasonable belief version of the reasons theory. In short, because I worry that the guiding reasons response collapses into the reasonable belief interpretation of the reasons theory, I am inclined to take seriously the first option, and to propose, albeit tentatively, that for the purposes of the reasonable belief account of justification an individual *X* acts on the basis of a (reasonable) belief *B* if (i) *B* is the sole reason for *X*’s acting, or (ii) *X* would not have acted if *X* had not believed *B*.

I therefore propose the following as a plausible supplementation of the reasons theory: *P* is justified in doing *A* in *C* if *P* has a reasonable belief that *A* is required to avoid harm to self in *C*, and *P* acts on the basis of that belief. Let us call this the *Reasonable Belief Theory*.

---

## IX A Practical Application

As I noted at the outset, Robinson suggests that one reason to prefer the deeds theory is that it “generates liability results that are more just and that better match our collective intuitions of what is just.”<sup>69</sup> To see whether this claim is plausible, let us apply what I am calling the Reasonable Belief Theory to some of the cases with which we began.

Consider again Mistaken Jogger. Suppose it is dusk, that Bert is running towards Alice quickly, and that he is waving his flashlight in a peculiar manner. And suppose that, because of Bert’s odd behavior, Alice forms the belief that she is about to be attacked. Then, I think, to the extent that Alice’s belief that she is about to be attacked is reasonable in the circumstances she would be justified in deploying self-defensive force. Reasonableness here has at least two components. First, it is surely the case that a reasonable person in Alice’s situation would conclude that the likelihood of attack is great. And second, surely Bert ought to realize that his actions might create a reasonable fear of attack in Alice. In terms of the allocation of risk, then, it seems right to say that Bert is imposing an unreasonable and so unjustifiable risk on Alice, and hence, that Alice is justified in using force to defend herself.<sup>70</sup> On the other hand, if it is only late afternoon, and Bert is merely jogging slowly, flashlight in hand, then it is not clear that Alice’s belief that she is about to be attacked would be a reasonable one in the circumstances. Since there is no reason to think that Bert is acting unreasonably or imposing an undue risk on Alice there is no reason for Alice to take defensive action. Moreover, were she to do so, she would lack a justifica-

tion even if she genuinely believed that an attack was imminent since, again, it is arguable that the belief upon which her defensive actions were based could not be held to be objectively reasonable.

Similarly, it seems clear that the Reasonable Belief Theory would regard Alice as having acted with justification in both Indecisive Attacker and Impossible Attacker. But in my view this is as it should be. For in Indecisive Attacker, risk is again allocated unequally and indifferently: because his behavior imposes undue risks on Alice and because he ought to know this, Bert cannot turn around and say that Alice may *not* take self-defensive actions on the grounds that, after all, he might not attack her. Given Bert’s unreasonable actions, Alice’s belief that she is about to be attacked is a reasonable one in the circumstances and is not affected by what Bert may eventually decide to do. The fact that in Impossible Attacker Bert will not succeed in attacking Alice is thus irrelevant to the question whether the risks are allocated appropriately as between Alice and Bert. With respect to the reasonableness of her justificatory belief, Alice’s epistemic situation in Impossible Attacker is no different from her situation in Genuine Attacker and Mistaken Jogger, and so ought to be treated in the same way in all three cases.

Of course, when faced with cases like Mistaken Jogger and Indecisive and Impossible Attacker it is clear what a deeds theorist will say. He will say that if Alice deploys self-defensive force on the basis of a reasonable belief that she is being attacked by Bert, then if Bert does not attack her or would not have attacked her, Al-

ice should be excused: she has acted badly, but understandably. In other words, a deeds theorist would agree with a reasons theorist in saying that Alice should not be punished in Mistaken Jogger and in Indecisive and Impossible Attacker, but would not infer from this that Alice's self-defensive actions were justified. In general, in cases in which reasons theorists see justifications, deeds theorists are likely to see excuses. How is one to decide between the two views?

Here I think two things are worth noting. First, to the extent that we are concerned with "our collective intuitions of what is just" I will simply note that I find it difficult to conclude that Alice acts badly in Impossible Attacker, which is what must be the case in order for Alice's actions to be excused rather than justified. Suppose we take the normative analysis of the distinction between justification and excuse and ask: what is it about Alice that we think results in her not being appropriately responsible for her actions? Perhaps we could say that it is her mistaken (albeit objectively reasonable) belief that she was about to be attacked that excuses her actions. Or perhaps we could say that she acted under something akin to duress, and that she should be excused for that reason. I do not doubt that we could say both these things. But again, both seem to get the phenomenology of the situation wrong. Alice acted as she should have acted and so should escape punishment as well as blame. To analyze Alice's culpability by reference to the concept of excuse is to analyze it by reference

to the theory of punishment and presuppose that she acts badly, and that, in my view, mislocates the crucial issue.

Of course, it might be objected that I and other reasons theorists are simply seeing what we want to see. For if I did not think Alice's actions in Impossible Attacker were permissible I would have no reason to prefer the Reasonable Belief Theory's analysis of that case to that of the deeds theory. But this brings me to the second point I wish to make, namely that the claim that Alice should be regarded as having acted with justification in each of Mistaken Jogger and Indecisive and Impossible Attacker is based on a certain theory about the relationship between justification and belief, and emphasizes the central role played by the concept of a wrong—as distinct from that of a harm—in the criminal law. To be sure, this will hardly impress Robinson, who is of the view that the criminal law is primarily concerned with the minimization of net overall harm. However, because that position is, in my view, problematic, reflection on the role played by the concept of a wrong in the criminal law provides reason to prefer the Reasonable Belief Theory's account of Mistaken Jogger and Indecisive and Impossible Attacker to that of the deeds theory.

Thus, contrary to what Robinson suggests, it seems to me that it is the Reasonable Belief Theory that "generates liability results that are more just and that better match our collective intuitions of what is just."

---

## X The Problem of Mistaken Justification

There is a final issue that needs to be addressed before our discussion of the Reasonable Belief Theory can be said to be complete and that is the problem of mistaken justification. In brief: how should the law treat individuals who act on the basis of a mistaken justificatory belief? Should such persons be regarded as having acted with justification, or should they merely be excused?

It is generally assumed that somebody who acts on the basis of a mistaken justification acts badly and should not be regarded as having acted with justification. And indeed, this is precisely what the deeds theory predicts. Says Robinson: "under the 'deeds' theory, a person who mistakenly believes that the conduct is justified is not justified. . . . Thus, the force used against the jogger-mistaken-for-an-attacker is not justified, although it may be excused if reasonable."<sup>71</sup> On this view, the reasonable-

ness of the belief is relevant to the question of punishment, but not to the question of whether the individual acted badly. Robinson's explanation for this should be familiar enough by now. Because what is important in determining the limits of justification is the world as it actually is and not the world as the actor takes it to be, a world in which an actor acts on the basis of a mistaken justification is a world in which that actor acts without justification. Depending on the nature of the mistake, of course, the law may choose to impose only partial liability and excuse the actor. But it remains the case that on the deeds theory, the law views the actor as having acted badly, and would caution others not to act in a similar manner.

But there are good reasons to question this account of mistaken justification. To see why, suppose that a police



officer arrests someone on suspicion of theft where all available evidence points to her guilt. It later turns out that she has been cleverly framed for the heist. Was her arrest justified or merely excused? It is typically assumed that in such a case the police officer acted with justification, indicating that “even within the law, no strict logic ties justification to actual facts rather than reasonably perceived facts.”<sup>72</sup> And indeed, one of the consequences of the Reasonable Belief Theory is that the notion of a mistaken justification is called into question. In saying this I do not mean to imply that Alice could not be mistaken about what Bert is about to do: clearly, Alice could think that Bert is about to attack her when he is simply out for a jog in the park. The point is rather that if Alice’s belief that justifying circumstances exist is reasonable, then the requisite justifying circumstances cannot fail to obtain. This is because the justifying circumstances are nothing over and above her reasonable belief that Bert is threatening her. In this sense, then, there cannot in general be mistaken justifications on the Reasonable Belief Theory. It may seem odd to say that the Reasonable Belief Theory precludes the general possibility of mistaken justifications. But consider: mistake as to justification arises on the Reasonable Belief Theory only in cases in which there is a mistake that goes to the *reasonableness* of the putative justificatory belief. Thus, if Alice unreasonably believes that Bert is about to attack her and acts on the basis of that belief, then she acts on the basis of a belief that cannot justify her actions and so acts badly. The difference as I see it is that with respect to mistaken justification the Reasonable Belief Theory draws the line in a different place than does the deeds theory. While the deeds theory views as mistaken any justificatory belief that fails to conform to the way things actually are, the Reasonable Belief Theory views as mistaken any justificatory belief that is unreasonable.<sup>73</sup> As Ripstein puts it,

Reasonableness is a description of the world from a particular perspective—the perspective of equality—and the requirement of belief is the further (and independent) requirement that those availing themselves of justifications know of the features of the situation that justify their acts. I’m not entitled to claim self-defense if I didn’t think you were attacking me. But reasonableness is an objective test of the adequacy of the belief: If the belief is reasonable, it can serve to justify the use of force, even if it turns out not to be true.<sup>74</sup>

This is why the Reasonable Belief Theory regards Alice as having acted with justification in Mistaken Jogger, In-

decisive Attacker, and Impossible Attacker. In all three cases Alice’s belief that she is about to be attacked is a reasonable one in the circumstances; thus, in all three cases Alice should be regarded as having acted with justification.

Can it be said that this claim “not only makes rights self-contradictorily subject to determination by the opinion of others; it also destroys the basis for rational adjudication?”<sup>75</sup> I do not see that it does. The Reasonable Belief Theory does not claim that my right to defend myself depends on the opinions or whims of others. Rather, the Reasonable Belief Theory claims that the limits of what is permissible are the limits of what is reasonable, where again, to act reasonably is to act out of respect for the equality of others. It is true that the concept of reasonableness as equality has an ineliminable objective or public component. But that simply means that to be reasonable a belief must be appropriately other-regarding, and must pay due attention to the interests of others. To insist on this does not destroy the basis for rational adjudication; to the contrary, to insist that justificatory beliefs be reasonable is precisely to insist that such beliefs be appropriate objects of rational reflection and scrutiny.

It turns out, then, that the expression “mistaken justification” is potentially ambiguous. On the deeds theory, an individual acts on the basis of a mistaken justification if the way the world is differs from the world as she takes it to be, and this is true even if she reasonably believes the world to be the way she takes it to be. According to the Reasonable Belief Theory, on the other hand, an individual acts on the basis of a mistaken justification if the belief on the basis of which she acts is not a belief on the basis of which she ought to have acted—if, that is, her justificatory belief is *unreasonable*. Both theories allow for the possibility of mistaken justification. However, in many cases of putative mistaken justification, such as Mistaken Jogger, in which the victim has a reasonable belief that she is being attacked, the Reasonable Belief Theory holds that there is no mistake as to justification and, hence, no need to revise our views about the liability results. Conversely, because according to the Reasonable Belief Theory an unreasonable belief cannot be a justificatory belief, an individual who acts on the basis of that belief will not be regarded as having acted with justification (and depending on the extent of the unreasonableness, such an individual might be denied an excuse as well).<sup>76</sup>

## XI Bystanders, Innocent Threats, and Innocent Aggressors

That being said, the Reasonable Belief Theory does allow that in certain other sorts of cases a reasonable belief may fail to justify. But the cases in which this occurs are puzzling ones. Call *X* an Innocent Aggressor if *X* is non-culpably aggressing against *Y*, perhaps under the influence of a powerful hallucinogenic drug administered without *X*'s knowledge or consent. Call *X* an Innocent Threat if *X*'s continued existence constitutes a threat to *Y*, but not because of anything that *X* has done.<sup>77</sup> And call *X* a Bystander if *X* does not herself endanger *Y*'s life and is not responsible for whatever it is that does endanger *Y*'s life.<sup>78</sup> How should the Reasonable Belief Theory deal with these sorts of cases?

It should be obvious that self-defense is impermissible when Bystanders are at issue (if indeed it makes sense to talk about self-defense in such situations). If a javelin is hurled towards *Y*, and *Y* reasonably believes that the only way to preserve his life is to grab an innocent stranger *X* and use him as a shield, then *Y* may not do so. The reason seems clear enough: because the Bystander is not responsible for whatever it is that is a threat to *Y*, *Y* may not prefer his life to *X*'s.

But what about Innocent Aggressors and Innocent Threats? These cases raise special problems. For if the reason why you may not sacrifice a Bystander to preserve your own life is that the Bystander is not responsible for the threat that endangers your life, then it would seem to follow that you may not deploy self-defensive force against Innocent Threats or Innocent Aggressors either since, again, *they* are not responsible for the threat that endangers your life. But if this is right, then the Reasonable Belief Theory would seem to be committed to the following: if Alice forms the reasonable belief that Bert is intent on killing her, and deploys self-defensive force on that basis, then she is to be regarded as having acted with justification even if it turns out that her belief was reasonable but mistaken, unless her mistake concerns Bert's agency or culpability, in which case she will at best be excused. And this is *prima facie* puzzling. For this amounts to the position that some mistakes do not matter for the purposes of justification, whereas others do.<sup>79</sup>

While I agree that this is puzzling, it seems to me that the best way to deal with these difficult cases is to distinguish them from cases of genuine self-defense: briefly, Innocent Aggressors and Innocent Threats should not be analyzed in terms of self-defense at all, but ought in-

stead to be analyzed as special cases of necessity. Moreover, because I believe necessity is best understood as an excuse, in my view somebody who kills an Innocent Threat or an Innocent Aggressor in order to preserve her own life ought to be excused for reasons of necessity, but should not be viewed as having acted with justification.<sup>80</sup> Although it may sound harsh, you act badly if you kill an Innocent Aggressor, even though your actions might be excusable in the circumstances. According to my interpretation of the Reasonable Belief Theory, then, the distinction between Innocent Aggressors and cases like Genuine Attacker has to do with the role played by agency and culpability: unlike a Genuine Attacker, an Innocent Threat does not act culpably because she does not *act*; and an Innocent Aggressor does not act culpably because, lacking responsibility, she is not appropriately *culpable* for her actions. And where culpable action is lacking, self-defensive action is unjustified, even if understandable and potentially excusable. This view is again based on a distinction between wronging and harming. Where an aggressor wrongs me by intentionally and culpably aggressing against me, I may with justification deploy self-defensive force. But when an aggressor merely harms or threatens to harm me, then even if it reasonably appears to me that I have been wronged, my self-defensive actions can only be excused.

To sum up: according to the Reasonable Belief Theory, in some situations a mistaken justification functions as a justification, whereas in other cases it merely functions as an excuse. Examples of the latter are situations where the mistake concerns the culpability or agency of the (apparent) threat. Does this count against the Reasonable Belief Theory? The deeds theory offers a unified account of mistaken justification: mistaken justifications excuse, but never justify. According to the Reasonable Belief Theory, on the other hand, most mistaken justifications justify, although some merely excuse. But is not a theory that offers a unified account of mistaken justification preferable to one that offers a disjunctive account? I do not believe that it is.

First, recall that on the Reasonable Belief Theory, mistaken justifications function as excuses only in cases like Innocent Aggressor and Innocent Threat, where culpability and agency are in issue. But these cases are extremely puzzling, and nobody yet has come up with a completely satisfactory analysis of them.<sup>81</sup> Second, the

problem of Innocent Aggressors and Innocent Threats is a problem for everybody. Consequently, the fact that the Reasonable Belief Theory has trouble with it does not count against it in particular. If a deeds theorist agrees that you act badly and without excuse if you sacrifice a Bystander to preserve your own life—as, presumably, he must—then how is he to avoid the conclusion that you act badly and without excuse if you sacrifice an Innocent Threat or an Innocent Aggressor to preserve

your own life? So while it must be granted that Innocent Aggressors and Innocent Threats create problems for the Reasonable Belief Theory's claim that mistaken justifications operate, in general, as justifications rather than as excuses, I doubt that such cases constitute a compelling reason to prefer the deeds theory to the Reasonable Belief Theory as a general account of the nature of justification.

## XII Conclusion

To conclude. Robinson's deeds theory of justification has a certain amount of intuitive force. The idea that objective circumstances are sometimes relevant to issues concerning liability makes some pre-theoretical sense and is supported by Fletcher's example of Williams, the failed bigamist. The problem is that this idea does not cohere with the way in which the criminal law polices the way individuals set the terms of their interactions with one another. I questioned Robinson's emphasis on harm and emphasized instead the role that the concept of wrongfulness ought to play in an account of justification, and in doing so I appealed to Ripstein's account of reasonableness as equality. This led me to conclude that Robinson's distinction between subjective "reasons" theories of justification and objective "deeds" theories of justification is problematic. I suggested two ways of understanding the relationship between justificatory beliefs and reasons for action, and indicated how I think the Reasonable Belief Theory ought to deal with

the problem of mistaken justification. In the end, I suggested the following: *P* is justified in doing *A* in *C* if *P* has a reasonable belief that *A* is required to avoid harm to self in *C*, and *P* acts on the basis of that belief.

Issues having to do with the nature of justification raise difficult problems, and there is certainly much more that could be said in defense of the deeds theory. Still, in my view the Reasonable Belief Theory, suitably supplemented with a notion of reasonableness as equality, is both theoretically attractive and intuitively plausible. To be sure, the extent of its attractiveness and plausibility will depend on one's views about the roles played by the concepts of wrongfulness and harm in the law. However, because I am of the view that the law is, and should be, primarily concerned with wrongs, I am also of the view that the Reasonable Belief Theory provides a better account of the nature of justification than does the deeds theory.

## NOTES

[Many thanks to Alan Brudner, John Kleinig, Arthur Ripstein, and, especially, two anonymous referees for this journal, for extremely helpful discussion and criticism.]

1 George Fletcher, *Rethinking Criminal Law* (New York: Oxford University Press, 2000), 683.

2 I take it that Joel Feinberg would endorse something along the lines of the first possibility. See, for example, Joel Feinberg, *The Moral Limits of the Criminal Law, Vol. I: Harm to Others* (New York: Oxford University Press, 1984). Judith Thomson, I think, would be happy enough with the second possibility. See Judith Jarvis Thomson, *The Realm of Rights* (Cambridge, MA: Harvard University Press, 1990). And Kant would presumably endorse something like the last possibility. See Immanuel Kant, *The Metaphysics of Morals*, trans. and ed. Mary Gregor (Cambridge: Cambridge University Press, 2003).

3 See J.L. Austin, "A Plea for Excuses" in *Philosophical Papers*, 3rd ed., ed. J.O. Urmson and G.J. Warnock (Oxford: Oxford University Press, 1979).

4 I take no stand on whether an agent claiming a justification is best interpreted as claiming that she has done nothing in which the law ought to take an interest, or as claiming that although she has committed a legally cognizable harm, she has done nothing wrong. For discussion of this distinction see George Fletcher, "The Nature of Justification" in *Action and Value in the Criminal Law*, ed. S. Shute, J. Gardner, and J. Horder (New York: Oxford University Press, 1993), 175.

5 Peter Westen, "An Attitudinal Theory of Excuse," *Law and Philosophy* 25 (2006): 289, 291.

6 Meir Dan-Cohen, "Decision Rules and Conduct Rules: On Acoustic Separation in Criminal Law," *Harvard Law Review* 97 (1984): 625.

7 Paul Robinson "Competing Theories of Justification: Deeds v. Reasons" in *Harm and Culpability*, ed. A.P. Simester and A.T.H. Smith (Oxford: Oxford University Press, 1996). Robinson first argued for a deeds theory of justification in Paul Robinson, "A Theory of Justification: Societal Harm as a Prerequisite for Criminal Liability," *UCLA Law Review* 23 (1975-76): 266. See also Paul Robinson, "Criminal Law Defenses: A Systematic Analysis," *Columbia Law Review* 82 (1982): 199; *Criminal Law Defenses* (St. Paul, MN: West Publishing Co., 1984); "A Functional Analysis of Criminal Law," *Northwestern University Law Review* 88 (1994): 857; *Structure and Function in Criminal Law* (Oxford: Clarendon Press, 1997); and "Justification Defenses in Situations of Unavoidable Uncertainty: A Reply to Professor Ferzan," *Law and Philosophy* 24 (2005): 775. For critical assessment of Robinson's view of justification see George Fletcher, "The Right Deed for the Wrong Reason: A Reply to Mr. Robinson," *UCLA Law Review* 23 (1975-76): 293; Brian Hogan, "The Dadson Principle," *Criminal Law Review* (1989): 679; and Michael Corrado, "Notes on the Structure of a Theory of Excuses," *Journal of Criminal Law and Criminology* 82 (1991): 465.

8 Robinson, "Deeds v. Reasons," 48.

9 Due to limitations of space, I am unable to discuss many interesting and important issues related to self-defense. Thus, I will not be concerned with issues about imminence, proportionality, or necessity. I will simply assume that the self-defensive actions in question are responses to imminent threats, and are necessary and proportional in relation to the nature and force of those threats. Similarly, for the most part I will only be concerned with cases in which one party *unlawfully* and *culpably* attacks another. In consequence, I will not discuss in much detail so-called Innocent Aggressors, Innocent Threats, or Bystanders. I am inclined to think that such cases ought to be treated differently, and that if self-defensive action is warranted in such cases it can only be because it is excused for reasons of necessity. I return to this issue below.

The literature on these sorts of cases, and the issues they raise, is voluminous. For a taste, see George Fletcher "Proportionality and the Psychotic Aggressor: A Vignette in Comparative Criminal Theory," *Israel Law Review* 8 (1973): 367 and "The Psychotic Aggressor: A Generation Later," *Israel Law Review* 27 (1993): 227; Cheyney Ryan, "Self-Defense, Pacifism, and the Possibility of Killing," *Ethics* 93 (1983): 508; Frances Kamm, "The Insanity Defense, Innocent Threats, and Limited Alternatives," *Criminal Justice Ethics* 6 (1987): 61; Judith Jarvis Thomson, "Self-Defense," *Philosophy & Public Affairs* 20 (1991): 283; Larry Alexander, "Self-Defense, Justification and Excuse," *Philosophy & Public Affairs* 22 (1993): 53 and "Justification and Innocent Attackers," *Wayne Law Review* 33 (1987): 1177; Suzanne Uniacke, *Permissible Killing: The Self-Defense Justification of Homicide* (Cambridge, UK: Cambridge University Press, 1994); Michael Otsuka, "Killing the Innocent in Self-Defense," *Philosophy & Public Affairs* 23 (1994): 74; Jeff McMahan, "Self-Defense and the Problem of the Innocent Attacker," *Ethics* 104 (1994): 252 and "Self-Defense and Culpability," *Law and Philosophy* 24 (2005): 751; Arthur Ripstein, "Self-Defense and Equal Protection," *University of Pittsburgh Law Review* 57 (1995): 685; Whitley Kaufman, "Is There a 'Right' to Self-Defense?" *Criminal Justice Ethics* 23 (2004): 20; Kimberly Kessler Ferzan, "Justifying Self-Defense," *Law and Philosophy* 24 (2005): 711; and Gerhard Øverland, "Self-Defence among Innocent People," *Journal of Moral Philosophy* 2 (2005): 127.

10 David Omerod, ed., *Smith & Hogan: Criminal Law*, 11th ed. (Oxford: Oxford University Press, 2005), 248. See also A.P. Simester and G.R. Sullivan, *Criminal Law: Theory and Doctrine*, 2nd ed. (Oxford: Hart Publishing, 2003), in which the authors choose not to classify defenses under the heads of justification and excuse.

11 See Fletcher, *Rethinking Criminal Law*, 760.

12 On excuse, see *Bourne* (1952) 36 Cr App R 125; *Cogan and Leak* [1976] QB 217, [1975] 2 All ER 1059; *United States v. Azadian*, 436 F.2d 81 (1971); and *State v. Haines*, 51 La. Ann. 731, 25 So. 372 (1899). On justification, see *Commonwealth v. Redline*, 391 Pa. 486, 137 A.2d 472 (1958).

13 David Horowitz, "Justification and Excuse in the Program of the Criminal Law," *Law & Contemporary Problems* 49 (1986): 109, 122.

14 For some reasons for supposing that the distinction might be less stable than one might think, see Kent Greenawalt, "The Perplexing Borders of Justification and Excuse," *Columbia Law Review* 84 (1984): 1897.

15 John Gardner, "The Gist of Excuses," *Buffalo Criminal Law Review* 1 (1997-98): 575, 590.

16 Kent Greenawalt, "Distinguishing Justifications from Excuses," *Law & Contemporary Problems* 49 (1986): 89, 108.

17 Austin, *A Plea for Excuses*, 177.

18 Wittgenstein asks: is it senseless to say: "Stand roughly there?" Ludwig Wittgenstein, *Philosophical Investigations*, 3rd ed., trans. G.E.M. Anscombe (Blackwell, 2003), §71.

19 Although the switch in terminology is potentially confusing, it is clear that Robinson means to equate reasons theories with subjectivist accounts of justification, and deeds theories with objectivist accounts of justification. Why the change in terminology? Robinson says: "I have substituted the 'reasons-deeds' terminology for the 'subjective-objective' because the latter has so many other uses with other meanings in other contexts that its use would seem to invite confusion." Robinson, "Competing Theories of Justification," 46, footnote 3. It remains to be seen whether Robinson's alternative terminology engenders any less confusion.

20 The example is taken from Robinson, "Competing Theories of Justification," 46.

21 Robinson, "Competing Theories of Justification," 46-7.

22 Robinson, "Competing Theories of Justification," 47.

23 The idea that issues having to do with harm are central to the criminal law is a familiar one from the work of Joel Feinberg. See Joel Feinberg, *Harm to Others*. For analysis of Feinberg's views, see Judith Jarvis Thomson, "Feinberg on Harm, Offense, and the Criminal Law: A Review Essay," *Philosophy & Public Affairs* 15 (1986): 381.

24 Robinson, "Competing Theories of Justification," 47.

25 The example comes from Simester and Sullivan, *Criminal Law*, 541.

26 Robinson, "A Theory of Justification," 287. Simester and Sullivan suggest that on Robinson's view Ashkenazi had a "right" to take the bag, "in that anyone who knew the true facts would not be allowed to interfere with Ashkenazi's tak-

ing of the bag." See Simester and Sullivan, *Criminal Law*, 541, note 29.

I am not sure what Robinson would say about similar cases. Take the following hypothetical: Bert wants Alice dead. He therefore puts a white powder—which he thinks is rat poison—in her coffee when she is not looking. Unbeknownst to Bert, however, Alice is suffering from a life-threatening illness and the white powder is the only antidote. Bert's actions therefore save Alice's life. Robinson would presumably say that Bert was justified in doing what he did, and that nobody who knew the facts of the situation would be allowed to interfere with Bert's actions. Would Robinson therefore conclude that Bert was not guilty of attempted murder? At one point Robinson seems to entertain the possibility that in such a case there might yet be "attempt liability." (See Robinson, "Competing Theories of Justification," 47.) But how could this be if, on his view, Bert's actions have positively beneficial consequences? For similar criticism see Simester and Sullivan, *Criminal Law*, 542, note 36.

27 Robinson, "Competing Theories of Justification," 63 (footnotes omitted).

28 For a nice discussion of this and other issues see Thomson, "Feinberg on Harm, Offense, and the Criminal Law."

29 I concede that this analysis is controversial since it might be objected that ignoring a person's wishes about what is to be done with her body—regardless of one's intentions and even regardless of the consequences—can never be a benefit to the person whose wishes were ignored. Although I disagree, I do not have the space to adequately address this issue here.

30 These examples are taken from Alan Brudner, "Agency and Welfare in the Penal Law" in *Action and Value in the Criminal Law*, ed. S. Shute, J. Gardner, & J. Horder (New York: Oxford University Press, 1993), 21.

31 Arthur Ripstein, "In Extremis," *Ohio State Journal of Criminal Law* 2 (2004-2005): 415, 417.

32 For an example supporting the proposition that the law is not concerned to protect merely harmful action, see *Fontainebleau Hotel Corp. v. Forty-Five Twenty-Five, Inc.*, 114 So. 2d 357 (Fla. Dist. CA 1959).

33 This observation is not limited to the criminal law. For it is likewise doubtful that harm is the linchpin upon which private law pivots. To see why, suppose that Alice, who is due to catch a plane to New York, is on her way to the airport, and that Bert, who is late for his plane, is driving carelessly. Bert hits Alice's car and injures Alice in the collision, and Alice misses her plane. The plane, however, crashes on take-off, killing everybody on board. Is Bert liable for his negligence? Surely he is. Has he *harmed* Alice? Arguably he has not. Granted, he injured Alice, but had he not collided with Alice, Alice would have died in the plane crash. To the contrary, then, Bert has positively benefited Alice, since the accident has resulted in an overall reduction in harm, at least where Alice is concerned. But this suggests that private law is not interested in harms as such; if it were, Bert ought to be praised for preventing a greater harm. Rather, just as the criminal law is plausibly viewed as being interested in the intentional violations of rights, private law is plausibly viewed as being interested in the certain kinds of rights infringements. This example is suggested by Ernest Weinrib, "Right and Advantage in Private Law," *Cardozo Law Review* 10 (1989): 1283.

34 George Fletcher, "The Right Deed for the Wrong Reason," 295.

35 *White v. State*, 44 Ohio App. 331, 185 N.E. 64 (1933).

36 Robinson, "Competing Theories of Justification," 46.

37 Robinson, "Competing Theories of Justification," 47.

38 Perhaps it will be objected that it is open to a deeds theorist to say that because Bert ought to know that his behavior will be interpreted by Alice as threatening, he is in fact threatening Alice, and hence that Alice is justified in defending herself in Indecisive Attacker. I am sympathetic to this view, and I will discuss it in detail below. For the moment, I simply note that whatever its merits it is not clear that it is something a friend of the deeds theory can help herself to, relying as it does on whether Bert's actions can be *interpreted by Alice* as being threatening.

39 *Brown v. United States*, 256 U.S. 335, 343 (1921).

40 See Arthur Ripstein, "Self-Defense and Equal Protection" and *Equality, Responsibility, and the Law* (Cambridge, UK: Cambridge University Press, 1999). For similar accounts see Gregory Keating, "Reasonableness and Rationality in Negligence Theory," *Stanford Law Review* 48 (1995-1996): 311, 323; Gregory Keating, "A Social Contract Conception of the Tort Law of Accidents" in *Philosophy and the Law of Torts*, ed. Gerald Postema (New York: Cambridge University Press, 2001), 22; and George Fletcher, "Fairness and Utility in Tort Theory," *Harvard Law Review* 85 (1972): 537.

41 For criticism, see, for example, John Gardner, "The Mysterious Case of the Reasonable Person," *University of Toronto Law Journal* 51 (2001): 273; Benjamin Zipursky, "Rawls in Tort Theory: Themes and Counter-Themes," *Fordham Law Review* 72 (2003-2004): 1923; and William Lucy, *Philosophy of Private Law* (New York: Oxford University Press, 2007).

42 See, for example, Lucy, *Philosophy of Private Law*, 406ff.

43 I have in mind here John Gardner. See John Gardner, "The Mysterious Case of the Reasonable Person."

44 See especially, John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993). See also W.M. Sibley, "The Rational Versus the Reasonable," *The Philosophical Review* 62 (1953): 554.

45 Rawls, *Political Liberalism*, 50.

46 Rawls, *Political Liberalism*, 51.

47 Rawls, *Political Liberalism*, 52.

48 Neil MacCormick, "Reasonableness and Objectivity," *Notre Dame Law Review* 74 (1998-1999): 1575, 1581.

49 Gary Schwartz, "Contributory and Comparative Negligence: A Reappraisal," *Yale Law Journal* 87 (1978): 697, 703.

50 Schwartz, "Contributory and Comparative Negligence."

51 Immanuel Kant, *The Metaphysics of Morals*, para. [6:231].

52 Keating, "Reasonableness and Rationality in Negligence Theory," 323.

53 Keating, "Reasonableness and Rationality in Negligence Theory," 327.

54 *Bolton v. Stone* [1951] AC 850 (HL). For an interesting dis-

discussion of what is involved in the creation and assumption of risk, see Andreas Teuber, "Justifying Risk," *Daedalus: Journal of the American Academy of Arts and Sciences* 119 (1990): 235.

55 159 F.2d 169 (2d Cir. 1947).

56 Ripstein, "Self-Defense and Equal Protection," 690.

57 The idea that one party cannot unilaterally and indifferently set the terms of her interactions with others is familiar from discussions of private law. See, for example, Ernest Weinrib, *The Idea of Private Law* (Cambridge, MA: Harvard University Press, 1995), Ripstein, *Equality, Responsibility, and the Law*, and Arthur Ripstein, "Philosophy of Tort Law" in *The Oxford Handbook of Jurisprudence and Legal Philosophy*, ed. J. Coleman and S. Shapiro (Oxford: Oxford University Press, 2001), 636.

58 Immanuel Kant, "On a Supposed Right to Lie from Benevolent Motives," *Grounding for the Metaphysics of Morals*, trans. James Ellington (Indianapolis: Hackett, 1993).

59 As an anonymous referee pointed out, analyzing wrongdoing solely in terms of risk allocation is controversial. If *X* assaults *Y*, then whatever else can be said about what *X* is doing, he cannot simply be said to be imposing a "risk" on *Y*. For that suggests that *X* is merely increasing the likelihood that something unfortunate will befall *Y* when what *X* is actually doing is making something unfortunate happen to *Y*. However, while wrongdoing surely goes beyond the mere imposition of risk, it is not implausible to say that the unfair imposition of risk is necessary for wrongdoing, even if it is not sufficient for it.

60 Perhaps sense could be made of the notion of bilaterality if it were framed in terms of consent. Thus, we could say that *X* and *Y* act bilaterally, or participate in a bilateral relationship, if *Y* consents to *X*'s assault, say. But while this might be plausible in some contexts—two hockey players who engage in a fight during the course of a game are not guilty of assault because they both consented to the fight—it cannot be correct in general, since consent does not vitiate criminal wrongdoing in all contexts: one cannot consent to be murdered.

61 Robinson, "Competing Theories of Justification," 50.

62 Robinson, "Competing Theories of Justification" 50.

63 Robinson, "Competing Theories of Justification," 50.

64 The criminal law is decidedly un-Kantian in this respect. That is, it does not matter *why* you do what you do so long as what you do is in accordance with the law. If I cross the street at the lights because that is what the law demands, that is all well and good. If I cross at the lights because I am convinced that, if I fail to do so, terrible things will happen to my pet dog, that is fine too. The criminal law is not concerned with what inclines or motivates us to obey the law; that we obey the law is enough. See, for example, Anthony Duff, "Rethinking Justifications," *Tulsa Law Review* 39 (2003-04): 829.

65 For a similar analysis of these sorts of issues, see Alon Harel, "Unconscionable Objection to Conscientious Objection: Notes on Sagi and Shapira," *Israel Law Review* 36 (2002): 219.

66 John Gardner, "Justifications and Reasons" in *Harm and Culpability*, ed. A.P. Simester and A.T.H. Smith (Oxford: Oxford University Press, 1996), 103.

67 John Gardner, "Justifications and Reasons," 105. Gardner

takes his inspiration from Joseph Raz's account of practical reasoning. See Joseph Raz, *Practical Reasons and Norms*, 2nd ed. (Oxford: Oxford University Press, 1999). For discussion and elaboration of these accounts of reasons and justifications, see Hamish Stewart, "The Role of Reasonableness in Self-Defense," *Canadian Journal of Law & Jurisprudence* 16 (2003): 317.

68 Gardner, "The Gist of Excuses," 579.

69 Robinson, "Competing Theories of Justification," 48.

70 I say that Bert ought to realize that his actions might lead Alice to form a reasonable belief that she is about to be attacked. But of course, whether Bert ought to know this will depend on what Bert knows about joggers and parks, and about what he knows about his own actions and behavior. However, this does not affect the point I am trying to make, which is that to the extent that reasonableness has a public component, what will count as a reasonable belief is determined in part by how suitably placed observers would view the situation. If Bert were to say, I did not know that people might think that jogging while waving a flashlight was threatening, it seems to me that his actions could still be viewed as being unreasonable, and that Alice's belief that she is about to be attacked could still be held to be reasonable in the circumstances.

71 Robinson, "Competing Theories of Justification," 47.

72 Kent Greenawalt, "Justifications, Excuses, and a Model Penal Code for Democratic Societies," *Criminal Justice Ethics* 17 (1998): 14, 23.

73 The etiology of the belief may matter here. If Alice forms the belief that Bert is about to attack her, that belief may be a reasonable one in the sense that a suitably placed observer would conclude that Bert's actions constitute a threat to Alice. Nonetheless, it might not constitute a justifying belief if Alice formed the belief on the basis of insufficient evidence. A friend of the Reasonable Belief Theory should therefore say that in order to be a justifying belief, the belief must be reasonable, and must also be formed in a reasonable way.

74 Ripstein, *Equality, Responsibility, and the Law*, 199.

75 Alan Brudner, "A Theory of Necessity," *Oxford Journal of Legal Studies* 7 (1987): 339, 364.

76 It should be clear that I part company with those who hold that an honest belief, even if unreasonable, is sufficient to negate culpability. Thus, in *DPP v. Morgan*, [1975] All ER 8 and *R. v. Pappajohn*, [1980] 2 S.C.R. 120, the accused argued that they lacked the requisite intent to commit rape due to their mistaken but honestly held belief that the victim had consented to sexual intercourse. In both cases it was held that an honest but unreasonable belief in consent was sufficient to negate mens rea. This might be thought to make trouble for the Reasonable Belief Theory since it suggests that the *reasonableness* of a mistaken belief is not necessarily relevant to the question of wrongdoing. In response, however, I would argue that in cases in which the risks and consequences of non-consent are acute, reasonableness has a vital role to play. Thus, I incline towards the view that in the case of rape, for example, belief in consent must be both honest and reasonable since the consequences of engaging in non-consensual sex are dire. Here I am in general agreement with the arguments of Edwin Curley, "Excusing Rape," *Philosophy and Public Affairs* 5 (1976): 325 and Toni Pickard, "Culpable Mistakes and Rape: Relating Mens Rea to

the Crime," *University of Toronto Law Journal* 30 (1980): 75.

77 The classic Innocent Threat is somebody who is pushed off the top of a cliff and whose falling body threatens another person who is sitting down below.

78 For discussion of these sorts of cases see Fletcher "Proportionality and the Psychotic Aggressor"; Thomson, "Self-Defense"; Otsuka, "Killing the Innocent in Self-Defense"; and McMahan, "Self-Defense and the Problem of the Innocent Attacker."

79 For discussion of this problem, see Ferzan, "Justifying Self-Defense"; Robinson, "Justification Defenses in Situations of Unavoidable Uncertainty"; and McMahan, "Self-defense and Culpability."

80 I acknowledge that this analysis is controversial, in part because the correct analysis of the defense of necessity is subject to debate. While some theorists view necessity as an excuse,

others argue that necessity operates instead as a justification. For an example of the former, see Ripstein, "In Extremis"; for an example of the latter, see Brudner, "A Theory of Necessity". Still, although I believe that necessity is best understood as an excuse, the important point for present purposes is simply that we should treat Innocent Threats and Innocent Aggressors differently than we should treat cases of culpable aggression, and this remains true even if one is of the view that necessity functions as a justification. Thus, even if one disagrees with my particular analysis of the defense of necessity, one can—and should—accept the more general proposition that Innocent Aggressors and Innocent Threats should not be analyzed in terms of self-defense at all, but ought instead to be analyzed as special cases of necessity.

81 Indeed, it may turn out that no satisfactory analysis is likely to be forthcoming. Jeff McMahan suggests this possibility in "Self-Defense and the Problem of the Innocent Attacker."