

“As If”: Reasonable Mistakes and Defensive Liability

Abstract

Jonathan Quong has recently developed a novel theory of the morality of defensive force, centered on the idea that one makes himself liable to defensive harm when and because they act *as if* other people lack moral rights they in fact possess. This essay interrogates this idea. I consider and reject Quong’s own development of the idea, before considering other ways of making the idea precise. In the end, I conclude that in as much as we have reason to endorse the idea that we make ourselves liable to defensive harm by treating others as if they lack rights they in fact possess, we actually have reason to reject Quong’s theory of defense in favor of one of its classic competitors.

1 Introduction

We typically wrong others if we harm them without their consent. One exception to this general rule, however, involves cases of self-defense or defense of others — cases where an aggressor *forfeits* some of his rights against harm.

Following the contemporary literature, let's say that someone who has forfeited some of their usual rights against defensive harm makes himself *liable* to harm. Someone is liable to some harm just in case they lack a right against that harm but not (exclusively) because they have consented to suffer it. The theory of liability to defensive harm has been a subject of great philosophical interest over the past few decades, and three accounts have emerged as most prominent.

First, there is Judith Jarvis Thomson's claim that a person is liable to (necessary and proportionate) defensive harm just in case they pose a threat of violating someone else's rights. On Thomson's view, posing such a threat is both necessary and sufficient for liability — no additional agency (e.g., fault) is required. Call this the **VIOLATION Account**.¹

Thomson's view has been widely rejected by contemporary theorists on the grounds that it makes liability "too easy" in that it does not require enough of a person's agency. For example, Kimberly Ferzan objects that on Thomson's account a person can lose their right to life as a result of bad brute luck, since one may be entirely blameless for posing a threat to another's rights.² Suppose, for example, that you've been duped into reasonably believing that I pose an imminent threat to your life. On Thomson's account, you make yourself liable to defensive harm if you attempt to harm me, even though you may be fully excused for doing so. But, worries Ferzan, "How can something that is sufficiently fundamental that we may kill to protect it, also be lost by mere misfortune such that another person has the right to kill us?"³ To eliminate the dependency of liability on luck, argues Ferzan, we need to supplement Thomson's posing-a-threat requirement with a culpability or fault requirement. This theory has come to be known as the **CULPABILITY Account**: a person is liable to (necessary and proportionate) defensive harm

¹Judith Thomson, "Self-Defense," *Philosophy & Public Affairs* 20, no. 4 (1991): 283-310.

²Kimberly Ferzan, "Culpable Aggression: The Basis for Moral Liability to Defensive Killing," *Ohio State Journal of Criminal Law* 9, no. 2 (2012): 669-698.

³Ferzan, "Culpable Aggression," 675.

just in case they *culpably* pose a threat of violating someone else's rights.⁴

Arguably the most popular theory, however, strikes many as a golden mean between VIOLATION and CULPABILITY — occupying a moderate position with respect to both the degree of agency required for liability and the extent to which liability is sensitive to luck. To illustrate, compare the following two cases:

Driver. You always keep your car well maintained and always drive carefully. Today, you decide to drive to the movie theater. On the way, a freak event that you could not have anticipated occurs that causes your car to veer out of control and towards someone who is sipping tea on their front porch. They will be killed unless someone kills you first.⁵

Phone. Your cell phone has, without your knowledge, suffered a wild and unforeseeable glitch such that when you hit the “send” button, the phone will transmit a signal that will detonate a bomb and kill an innocent person. Someone who is made aware of the glitch hurries to intervene. Since you have earbuds in and are enjoying your music turned up loud, they cannot warn you and can prevent you from pressing the send button only by killing you.⁶

According to VIOLATION, you are liable to be harmed in both cases.⁷ According to CULPABILITY, you are liable in neither. Yet there may seem a morally

⁴See Ferzan, “Culpable Aggression,” and “Justifying Self-Defense,” *Law and Philosophy* 24, no. 6 (2005): 711-749; (early) Jeff McMahan, “Self-Defense and the Problem of the Innocent Attacker,” *Ethics* 104, no. 2 (1994): 252-290; Phillip Montague, “Self-Defense and Choosing Between Lives,” *Philosophical Studies* 40, no. 2 (1981): 207-219. In the neighborhood is Susanne Burri’s “Subjective Impermissibility” account in her “Defensive Liability,” *Criminal Law and Philosophy* 16, no. 3 (2012): 539-553.

⁵Jeff McMahan, “The Basis of Moral Liability to Defensive Killing,” *Philosophical Issues* 15, no. 1 (2005): 386-405; 393

⁶*Ibid.*, 387.

⁷Or so thought Thomson, and so have assumed many of her critics. Note that VIOLATION

relevant difference between the two cases: only in the first does your action make the resulting threat a *reasonably foreseeable* consequence of that type of action.⁸ For this reason, argues McMahan, only in Driver are you *morally responsible* for the outcome: the harm in Driver is a matter of bad option luck, whereas the harm in Phone is a matter of bad brute luck.⁹ This may seem a crucial difference when it comes to requiring a person to internalize costs they have externalized onto others. A person seems to have no standing to complain if they are made to bear the costs of a gamble they knowingly undertook, even if that gamble was permissible given their evidence.

These luck-egalitarian considerations incline many theorists to endorse a theory that assigns liability in Driver but not in Phone — the so-called **RESPONSIBILITY Account**. According to this account, a person is liable to (necessary and proportionate) defensive harm just in case they are *morally responsible* for posing a threat of violating someone else's rights.¹⁰

These three accounts have largely dominated the literature. And so it is re-

implies liability in Phone only on the assumption that pressing the “send” button would violate the innocent victim's rights. This assumption is not obviously true.

⁸Jeff McMahan claims that some unjust harm is a reasonably foreseeable outcome of an action ϕ in the relevant sense just in case either (i) ϕ is a token of a risk-imposing type of activity, or (ii) in the circumstances the agent ought to have foreseen [that ϕ -ing] carried a non-negligible risk of causing a significant unjust harm” (*Ibid.*, 397).

⁹An outcome is a matter of brute luck for you to the extent it results from factors beyond your control, and a matter of option luck to the extent it results from factors within your control — for instance, as a foreseeable consequence of a deliberate choice. In Ronald Dworkin's words, option luck “is a matter of how deliberate and calculated gambles turn out — whether someone gains or loses through accepting an isolated risk he or she should have anticipated and might have declined.” See Ronald Dworkin, *Sovereign Virtue* (Oxford, 2000), 73.

¹⁰Theorists who endorse some version of this account include: Helen Frowe, *Defensive Killing* (Oxford, 2014), 72–87; Kerah Gordon-Solmon, “What Makes a Person Liable to Defensive Harm?,” *Philosophy and Phenomenological Research* 97, no. 3 (2018): 543–67; Michael Otsuka, “Killing the Innocent in Self-Defense,” *Philosophy and Public Affairs* 23, no. 1 (1994): 74–94; Jeff McMahan, “The Basis,” and *Killing in War* (Oxford, 2009), 155–202; and Michael Otsuka, “The Moral Responsibility Account of Liability to Defensive Killing,” in *The Ethics of Self-Defense*, ed. Christian Coons and Michael Weber (Oxford, 2016), 51–68. Not all proponents of this account agree with McMahan as to how best to understand the relevant notion of “moral responsibility,” though they all agree that moral responsibility does not imply culpability.

refreshing to see a novel theory enter the mix of late — a theory that, like RESPONSIBILITY, attempts to occupy a middle ground between VIOLATION and CULPABILITY both in terms of the agency it requires for liability and in terms of the extent to which it makes liability sensitive to luck. This is Jonathan Quong’s “Moral Status” theory of liability.¹¹ Quong agrees with McMahan and other defenders of RESPONSIBILITY that culpability is not required for liability (contra CULPABILITY), that merely posing a threat of violating a right is insufficient for liability (contra VIOLATION), and that the extra agential ingredient needed has to do with engaging in certain kinds of risk-imposing behavior. Quong rejects, however, the claim that moral responsibility is that extra agential ingredient. Indeed, on Quong’s view one can foreseeably bring about a threat to another’s rights without incurring liability. On Quong’s view what matters, rather, is whether one makes a certain kind of mistake about the rights that others have. What matters is whether (in a special sense that we’ll unpack shortly) one treats others “as if” they lack rights they in fact have.

This essay interrogates this idea that liability is grounded in the phenomenon of treating others “as if” they lack certain rights. In §2 I unpack Quong’s particular precisification of this idea and his motivation for endorsing it. In §3 I express three worries for Quong’s account. In §4 I abstract away from Quong’s particular account to determine the most plausible version of the idea that liability is grounded in treating others “as if” they lack certain rights. I argue that, in the end, an appeal to the as-if phenomenon probably does more to motivate a theory of liability like Thomson’s VIOLATION than Quong’s MORAL STATUS.

2 The Moral Status Account

The idea at underlying Quong’s account is a simple one. It’s an idea he describes as “independently plausible”:¹²

¹¹Jonathan Quong, *The Morality of Defensive Force* (Oxford, 2020), ch. 2; and Quong, “Liability to Defensive Harm,” *Philosophy & Public Affairs* 40, no. 1 (2012): 45-77.

¹²*Defensive Force* : 39.

The intuitive idea is that when we act *as if* other people lack moral rights against the harms that we might impose, but those others possess these moral rights, we lose the standing to complain if defensive harm is imposed on us to avert our wrongful threats.¹³

I'll call this Quong's

Ur-Idea. We make ourselves liable to defensive harm when and because we mistakenly act as if other people lack the moral rights that are threatened by our action.

As we'll see later, "acting as if someone lacks rights they in fact possess" is ambiguous. As a result, there are different ways of fleshing out the *Ur-Idea*. Quong offers one particular way of doing so. Understanding his specification requires understanding, first, the difference between so-called *fact-relative* and *evidence-relative* moral permissibility.¹⁴ As Quong uses terms, an act is permissible in the fact-relative sense if it would be permissible "if we knew all the relevant facts," and permissible in the evidence-relative sense if it would be permissible "if the relevant facts were what the available evidence gives us sufficient (or apparently sufficient) reason to believe they are."¹⁵ To illustrate the difference, suppose an unconscious patient is brought to the emergency room, and all the available evidence gives the doctor compelling reason to believe the patient needs his arm immediately amputated. The evidence, however, is misleading, and as a matter of fact the amputation will in no way benefit the patient. In this case, the doctor would have an evidence-relative, but not a fact-relative, permission to amputate the leg.

¹³Jonathan Quong, "The Morality of Defensive Force: Replies to Christie, Hecht, and Parry," *Journal of Moral Philosophy* 20, no. 5-6 (2023): 462.

¹⁴To my knowledge, this language traces back to Derek Parfit, *On What Matters*, vol. 1 (Oxford, 2011), 150-151.

¹⁵*Defensive Force*, 20.

The particular form of treating another “as if” they lack rights that Quong is interested in is one that he characterizes in terms of evidence-relative permissibility:¹⁶

MORAL STATUS Account. A is liable to defensive harm for ϕ -ing just in case (i) there is some right R had by someone V such that by ϕ -ing A poses a threat (or apparent threat) to R, and either (ii) A *lacks* an *evidence-relative* permission to ϕ , or (iii) A *has* an evidence-relative permission to ϕ but *only because* A’s evidence gives him sufficient reason to believe that V lacks right R.¹⁷

This account posits two paths to liability. The first path is captured by the conjunction of (i) and (ii): I make myself liable to defensive harm if I *threaten your right by engaging in evidence-relative impermissible conduct*. VIOLATION shares this commitment. And it is a claim that RESPONSIBILITY and CULPABILITY are *defeasibly* committed to: on those theories, posing a threat to your rights by way of evidence-relative impermissible conduct renders me liable so long as there is not a defeater to my responsibility or fault (such as duress).

The really innovative feature of MORAL STATUS concerns condition (iii). Quong proposes a novel, non-fault-based path to liability, captured by the conjunction of conditions (i) and (iii). A person who poses a threat as a result of evidence-relative *permissible* conduct is liable just in case *the evidence-relative permissibility of their conduct “depends on” the mistaken assumption that the victim lacks a certain right*. Says Quong:

¹⁶*Ibid.*, 38; Quong, “Replies,” 464–470.

¹⁷Quong sometimes formulates the account in conditional, rather than biconditional form, in a way that suggests the account provides merely a *sufficient* condition for liability (*Defensive Force*, 38). Elsewhere he formulates the account in biconditional form (*Defensive Force*, 41), suggesting both necessary and sufficient conditions. The biconditional formulation is clearly the intended one, as many of the implications Quong claims for the account require the necessity claim. His verdict that the conscientious driver is *not* liable, for example, is based on the fact that the evidence-relative permissibility of his driving does not depend on mistaken evidence about the pedestrian’s rights. This inference is unavailable if the account states only a sufficient condition for liability.

Even when we do not act wrongly in the evidence-relative sense, we sometimes gamble with the moral rights of others in the sense described above: we act in ways that we ought to know would be wrong *but for the assumption that certain others lack rights*. Though it does not make sense to blame us when we act in these ways, it does make sense to say that we may forfeit some of our rights when we take these sorts of risks.¹⁸

It is helpful to illustrate the idea with a case. Consider:

Resident. The identical twin of a notorious mass murderer is driving in the middle of a stormy night in a remote area when his car breaks down. He is not in a position to suspect that his twin brother, the murderer, has within the past few hours escaped from prison in just this area, and that the residents have been warned of the escape. The murderer's notoriety derives from his invariable *modus operandi*: he violently breaks into people's homes and kills them instantly. As the twin whose car has broken down approaches a house to request to use the telephone, the resident of the house takes aim to shoot him, reasonably believing him to be the murderer and her life to be in imminent danger.¹⁹

The resident is a non-culpable aggressor, since (by stipulation) her evidence that the twin poses an imminent threat is compelling enough to give her an evidence-relative permission for attacking him. But the resident has this permission *only because* her evidence is mistaken about the twin's rights, as a result of her being mistaken about the twin's identity. If her evidence didn't support the belief that the twin was liable to be harmed, then she wouldn't have an evidence-relative permission for attacking him. This is why, according to

¹⁸Quong, *Defensive Force*, 37–38 (italics mine).

¹⁹The case, I believe, originates with Jeff McMahan, "The Basis," 387, and appears in Quong, *Defensive Force*, 24.

MORAL STATUS, she is liable to be harmed even though she is non-culpable. (Quong's 'but for' language in the quote above suggests a counterfactual understanding of dependence: the permission depends on a rights mistake just in case the agent would lack the permission if their evidence didn't contain that mistake.)

Like RESPONSIBILITY, MORAL STATUS occupies a moderate position between VIOLATION and CULPABILITY. Contra VIOLATION, more is required than merely posing a threat to someone's rights; one must pose a threat in a particular kind of way. Contra CULPABILITY, fault isn't required. But there are important differences between RESPONSIBILITY and MORAL STATUS. Most notably, according to MORAL STATUS one can fail to be liable even when one poses a foreseeable threat.

Recall the case of Driver, for instance, where despite your conscientious driving you pose a threat to an innocent person's rights. According to McMahan, the relevant test for liability is a foreseeability test. Is it reasonably foreseeable that taking a car on the road might result in a threat to an innocent person? The answer is clearly Yes:

What makes [the driver] liable is that, as a morally responsible agent, he voluntarily chose to set a couple of tons of steel rolling as a means of pursuing his ends, knowing that this would involve a tiny risk that he would lose control of this dangerous object that he had set in motion, thereby imperiling the lives of the innocent...the driver is liable because he voluntarily engaged in a risk imposing activity and is responsible for the consequences when the risks he imposed eventuate in harms.²⁰

MORAL STATUS utilizes a different test. Instead of asking whether the threat was foreseeable, we ask, first, whether you (the driver) had an evidence-relative permission for taking your car out on the road, and, second, whether

²⁰McMahan, "Moral Basis," 394.

that permission depends on your evidence failing to attribute rights to the would-be victim that she in fact possesses. The answers, respectively, are Yes and No: “The evidence-relative permissibility of getting behind the wheel of a vehicle and imposing a tiny risk of harm on others does not depend on treating those others as if they lack any rights against harm.”²¹ This is because, says Quong:

In the case of prudent driving ... the fact that [the driver] imposes a tiny risk of harm on others is consistent with everyone’s moral status, since the risk of harm is (we are assuming) justifiable to all persons in light of the balance of costs and benefits of the practice of prudent driving.²²

For this reason, MORAL STATUS gives a different verdict from RESPONSIBILITY in the case of Driver and declares the driver *non*-liable.

Quong takes this to be a virtue of his account. He finds it extremely counterintuitive that the driver is liable to defensive harm.²³ In Quong’s view, MORAL STATUS strikes a sweeter spot than RESPONSIBILITY in the way it discriminates between different cases of non-culpable aggression. In contrast to RESPONSIBILITY, MORAL STATUS delivers asymmetric verdicts in Resident and Driver: the non-culpable threatener is liable in Resident but not in Driver. Quong finds this to be the intuitively correct result.

3 Worries for the Moral Status Account

The MORAL STATUS account is fresh and ingenious. Nonetheless, I doubt it plausibly captures the way in which acting “as if” others lack rights is relevant to liability. I have three worries.

²¹Quong, *Defensive Force*, 39.

²²*Ibid.*, 37.

²³*Ibid.*, 39. Though Quong believes that the pedestrian would nonetheless be permitted to harm the driver in self-defense. On his view, the pedestrian’s agent-relative prerogatives permit her to infringe the driver’s rights to defend herself. See: *Ibid.*, 58–95.

My first concern is that MORAL STATUS is too narrow in its focus on the kinds of mistakes that can give rise to liability. It claims that evidence-relative-permissible uses of force generate liability only when the permission depends on a mistake about *what rights their target has*. Resident is an example where the permission depends on such a mistake. So too are cases where one makes no mistake about the identity of their target, but has mistaken evidence that their target has waived or forfeited rights. Let's call these *mistakes of rights attribution*: these are mistakes where one believes, of her target, that he fails to possess rights against whatever harm her action imposes on him.

The problem for MORAL STATUS is that there are other kinds of mistakes that also seem to give rise to liability. Consider, first:

Scare the Bear. Hiker is hiking in the wilderness, believing himself to be alone. He stumbles across a hungry grizzly bear that gives chase. Hiker has good reason to believe that he can defend his life only by throwing a grenade in the bear's direction to scare it off. Hiker is quite confident, though not sure, that there is no one in the vicinity of where he is throwing the grenade. Unfortunately, two of his neighbors, Alice and Bob, happen to be just out of sight and exactly where Hiker throws the grenade. They can defend themselves only by kicking the grenade back to Hiker, harming him.

I find it intuitively compelling that Hiker makes himself liable to have the grenade kicked back, despite the fact that he has an evidence-relative permission to throw the grenade. Cases like these make trouble for CULPABILITY, since Hiker is not culpable for the threat he poses to Alice and Bob. Hiker is not culpable because he has very good reason to believe that throwing the grenade is necessary to save his life, and very little reason to believe that anyone else is in the vicinity. (By contrast, VIOLATION and RESPONSIBILITY both imply that he *is* liable.) But this case also makes trouble for MORAL STATUS, as

this theory also implies that Hiker is not liable. This is because Hiker does not make any mistakes about people's rights: his evidence (we can stipulate) gets it exactly right when it comes to the question whether Alice and Bob have forfeited rights. He correctly believes that they are in possession of all their usual rights to not be harmed. This, then, is not a case involving a mistake of rights attribution, by contrast to a case like Resident. Hiker's mistake is only a mistake about whether anyone will be harmed by his action.

But isn't it just as morally serious for the evidence-relative permissibility of one's conduct to depend on the accuracy of one's evidence about *whether rights are put at risk* by that conduct as for it to depend on the accuracy of one's evidence about *who has what rights*? Moreover, isn't tossing the grenade on the basis of this kind of mistake *also* a way of "gambling" with people's rights?²⁴ The gamble in this case is just the gamble that there is no one in the vicinity of the grenade — a gamble one surely takes whenever one throws an inherently dangerous explosive device into an area where they know there is some chance that other people are present. So the worry for MORAL STATUS is that it is unjustifiably focused on only one kind of mistake about, and way of gambling with, the rights of others.

There are yet further kinds of mistakes that give rise to liability. Consider:

Warehouse. Adam is informed by many extremely reliable sources that a villain is holed up in a long-since-abandoned warehouse, and is preparing to use a remote detonator to set off a bomb that will kill thousands of people. Adam's sources inform him that the

²⁴This language is a call back to the Quong passage we quoted earlier (Quong, *Defensive Force*, 37):

Even when we do not act wrongly in the evidence-relative sense, we sometimes gamble with the moral rights of others ... we act in ways that we ought to know would be wrong but for the assumption that certain others lack rights... It does make sense to say that we may forfeit some of our rights when we take these sorts of risks.

detonation can be prevented only if Adam throws a grenade into the warehouse, killing the villain and seriously injuring two innocent bystanders, Beatrice and Charlie, who have been taken hostage. Adam throws the grenade, knowing that this will cause Beatrice and Charlie's rights to be infringed, but believing that the infringement is permissible in light of the thousands of lives that will be saved. As it turns out, there is no villain in the warehouse — only Beatrice and Charlie. They will be seriously harmed unless they kick the grenade back towards Adam, in which case he will be seriously harmed instead.

It is very compelling that Adam makes himself liable to be harmed. But MORAL STATUS implies otherwise. Adam does not make any mistake of rights attribution.

But he also doesn't make the same sort of mistake as in Scare the Bear. Adam correctly judges who will be harmed by his conduct and whose rights infringed. He makes no mistakes about whether anyone will be harmed by his action; he makes no mistakes about the identity of those who will be harmed by his action; and he makes no mistakes about what rights those people have. His only mistake is about whether the rights and interests of third-parties provide sufficient reason to *justify* the infringement of Beatrice and Charlie's rights. His evidence gives him sufficient reason to believe that there is such a justification, when in fact there is not. But again: isn't tossing a grenade into the warehouse on the basis of *this* kind of mistake just as much a way of "gambling" with people's rights as doing so on the basis of mistakes of rights attribution?

So the first worry for MORAL STATUS is this. The theory claims that someone who imposes an evidence-relative-permissible threat to others makes himself liable only if he makes what I've called a *mistake of rights attribution*: he mistakenly believes of those whom he threatens to harm that they lack rights against that harm. The worry for MORAL STATUS is that it seems evidence-

relative-permissible threateners can make themselves liable by way of other kinds of mistakes, such as mistakes about whether anyone would be harmed at all (Scare the Bear), or whether there is a justification for infringing the rights of those who would be harmed (Warehouse).

A second worry for MORAL STATUS is that it discriminates between cases where it seems it shouldn't. Consider the following pair of cases. (Stipulate, for the sake of illustration, that (i) it is fact-relative permissible to redirect a trolley towards one innocent person to save five, (ii) it is fact-relative impermissible to redirect towards one innocent person to save *fewer* than five, and (iii) it is fact-relative permissible to redirect towards one *liable* person to save even just one innocent person.)²⁵

Waiting on Trolley 1. Alice sees a runaway trolley barreling down the tracks. Her evidence is that there are five innocent people stuck on the primary track, out of eyesight. Alice can redirect the trolley at t_1 , in which case it will break Brianna's legs, whom Alice has good reason to think is the person at fault for the runaway trolley. In fact, there is no one threatened by the runaway trolley on the primary track, and Brianna had nothing to do with the runaway trolley. Alice redirects the trolley. Brianna can defend herself only by breaking Alice's legs.²⁶

²⁵If you disagree with those particular numbers, feel free to adjust to taste.

²⁶This case has similarities to a case from Lars Christie in his critique of MORAL STATUS in his "Mortal Mistakes," *Journal of Moral Philosophy* 125, no. 1 (2022): 9. Christie's case is also one where an agent has an evidence-relative permission to turn the trolley on account of evidence that doing so will save a number of innocent people (though in his version, the agent does not have any evidence that their target lacks rights, and they turn the trolley with the intention of murdering their target). Christie and I deploy our cases in service of different objections, however. Christie argues that MORAL STATUS is too narrow because it fails to imply liability in some cases involving culpable threats. My objection is that Quong's formulation of his account in terms of evidential "dependence" generates arbitrary asymmetries. The key feature of my cases is that they're presented as a comparative pair: Alice makes the very same mistake about Brianna's rights in both cases, acts on the same evidence and for the same reasons, and yet MORAL STATUS delivers different verdicts based solely on whether Alice happens to have a counterfactual backup justification she's unaware of. This isn't at all

Waiting on Trolley 2. As before, except that Alice does not redirect the trolley at t_1 . At t_2 she receives a radio call that changes her evidence. Her evidence is now that there are only two innocent people stuck on the tracks. (As before, there is in fact no one on the primary track.) She still has opportunity to redirect the trolley towards Brianna, whom she still has good reason to think is the person at fault for the runaway trolley. She redirects the trolley. As before, Brianna is innocent and can defend herself only by breaking Alice's legs.

In both cases Alice has an evidence-relative, but not a fact-relative, permission to redirect the trolley. She has an evidence-relative permission in both cases because in both cases she has excellent evidence that Brianna is liable to be harmed. She lacks a fact-relative permission in both cases because in both cases Brianna is not liable to be harmed and there is no one who benefits from harming her. The only difference between the two cases concerns Alice's evidence about the number of people on the primary track.

I find it very hard to believe that *this* difference could make a difference as to whether Alice is liable to be harmed. Surely, if Alice is (not) liable to be harmed in the one case, she is (not) liable to be harmed in the other.

BUT MORAL STATUS seems to deliver asymmetric verdict in these two trolley cases. According to MORAL STATUS, Alice is liable just in case her evidence-relative permission *depends on* mistaken evidence about Brianna's rights, where Quong cashes out this dependence in counterfactual, "but for" terms.²⁷ Only in the second case, however, do we find such dependence. In the first case,

about culpability, but about whether Quong's account tracks morally-relevant features of one's evidence, and about whether his account plausibly captures the idea of acting "as if" someone lacks rights.

²⁷“Even when we do not act wrongly in the evidence-relative sense, we sometimes gamble with the moral rights of others in the sense [that] we act in ways that we ought to know would be wrong *but for* the assumption that certain others lack rights” (Quong, *Defensive Force*: 37-38).

Alice's evidence-relative permission is not dependent on her mistake about Brianna's liability. This is because that permission is "backed up" by a lesser-evil justification: Alice would have an evidence-relative permission to redirect the trolley even if she knew that Brianna was innocent, since, by the lights of her evidence, she would still have a lesser-evil justification for redirecting the trolley. In the second case, by contrast, Alice's permission isn't backed up by such a lesser-evil justification. In that case, then, her permission *does* depend on his mistake about Brianna's rights. MORAL STATUS thus implies that Alice makes herself liable in the second case, but not in the first.²⁸

This implication is very counterintuitive on its face. And it looks even more counterintuitive the closer we squint at it. In both trolley cases Alice makes the very same kind of mistake about what rights Brianna has. The difference

²⁸This objection is predicated on Quong's counterfactual gloss on 'depends on,' according to which X's evidence-relative permission to ϕ depends on X's evidence e just in case X would lack that permission if he were not in possession of e . This is the reading best supported by Quong's use of "but for" language. We might wonder, however, whether the trolley objection can be diffused by opting for a different notion of dependence.

I'm dubious. A *causal* reading won't do. Evidence-relative permissions, as Quong defines them, are themselves counterfactual in nature; they are determined by asking what *would* be permissible if the facts were as one's evidence suggests. The answer to this counterfactual question is determined by one's evidence together with the correct moral principles. On this picture, there's no sense to be made of the idea that a bit of evidence *causes* an evidence-relative permission. Evidence explains the permission, but not by causing it.

The only alternative I see to a counterfactual or causal notion of dependence is an *explanatory* one, according to which the evidence-relative permissibility of ϕ -ing depends on evidence e just in case e is part of the set of evidence *in virtue of which* ϕ -ing is evidence-relative permissible. But what is that body of evidence in virtue of which something is permissible? It's hard to resist the answer that it is one's *total* body of evidence, since *what it is* for an action to be evidence-relative permissible is for the facts to be as one's total evidence gives one sufficient reason to believe they are. But MORAL STATUS would make liability contingent on intuitively irrelevant kinds of mistakes if any mistaken evidence in one's total body of evidence about people's rights could make one liable. For example, imagine the case of Driver. On this total-evidence form of explanatory dependence, the conscientious driver could be liable if he had mistaken evidence that, for example, the pedestrian has consented to an unrelated medical procedure that she hadn't, or that she had forfeited rights against punishment by committing some crime the previous week.

We might try to resist such promiscuity by arguing that the evidence in virtue of which some action is permissible isn't one's total body of evidence, but rather some *relevant* subset of that evidence. But how else to distinguish between "relevant" and "irrelevant" evidence without falling back on counterfactual dependence?

between the cases, again, is just a difference about whether Alice's evidence-relative permission depends on this mistake. But why should the dependence relationship between the mistake and the evidence-relative permission determine liability, rather than the mistake itself?

Moreover, it's not just that Alice makes the same mistake in both cases; it's also that she intuitively "gambles" with Brianna's rights in both cases. She risks being wrong, not just about what rights Brianna has, but also about whether the rights of others justify the infringement of Brianna's rights. So the question now becomes: Why should Quong's particular way of characterizing the phenomenon of gambling with the rights of others determine liability rather than the more familiar kind of gamble that Alice's makes in *both* cases?

So it's not just that MORAL STATUS has counterintuitive implications. My third and final worry for the account that cases like Waiting on Trolley bring out is that MORAL STATUS looks inadequate as a precisification of the idea that centrally motivates the account: Quong's *Ur-Idea*. Step back a moment and consider where we've ended up in relation to that idea. The idea, recall, is that we make ourselves liable to defensive harm when and because we mistakenly act *as if* other people lack the moral rights that are in fact threatened by our action. Ignore, for the moment, the question whether it is plausible that acting in this way grounds liability. Consider the more basic question *what it means to act in this way*. One worry that emerges from reflection on our cases is that Quong's answer to this latter question is dubious in its own right. Quong is offering an account of a pre-theoretic concept; the idea of "acting as if others lack rights" is not a term of art. But, for my part, I can't detect any sense of the "as if" language where it seems felicitous to say, for example, that Alice treats Brianna as if she lacks rights in one of the trolley cases but not in the other. This leads me to worry whether Quong's characterization of the "as-if" phenomenon is adequate in its own right — independent of the theory of liability he builds on top of it.

In summary, we can distinguish three worries for MORAL STATUS. First, it allows liability to arise on the basis of an implausibly narrow range of mistakes we might make. It predicates liability on one kind of mistake (mistakes of rights attribution) while ignoring other kinds of mistakes that can give rise to liability. Second, it makes liability dependent on factors that seem morally irrelevant, such as whether or not one's evidence-relative permission is overdetermined, as illustrated in the pair of Waiting on Trolley cases. Both of these worries constitute a comparative disadvantage for the MORAL STATUS, since two of its competitors — VIOLATION and RESPONSIBILITY — deliver intuitively correct results in all of the above cases.

The third worry is that MORAL STATUS offers a descriptively implausible account of what it is to treat someone as if they lack rights. It offers a characterization of the as-if phenomenon that does not seem to capture anything we would pre-theoretically recognize as a kind of that phenomenon. For this reason, MORAL STATUS fails to be a precisification of Quong's *Ur-Idea*, and thus fails to be motivated by the plausibility of that idea.

4 Clarifying “As If”

So: can we flesh out this notion of *treating someone as if they lack rights they in fact possess* in a way that is both independently plausible and that explains why the mistakes made by the aggressors in Scare the Bear, Warehouse, and Waiting on Trolley make those aggressors liable to defensive force?

To answer the question, we first need to disambiguate the locution, “treating X as if she lacks rights.” There is not a univocal phenomenon that is picked out by this language. Plausibly, one thing this language might pick out is a *deliberative* phenomenon. In this sense, to treat someone as if they lack a certain right is to fail to appropriately take that right into consideration in one's practical deliberations. There are two different ways one might fail to appropriately take a right into consideration. The first is an evidence-relative, deliberative failing: one fails to give someone's right the weight in one's deliberations that one ought to given one's evidence. The second is a

fact-relative, deliberative failing: one fails to give someone's right the weight in one's deliberations that one ought to given the facts.

Acting on the basis of such deliberative failings is one way of acting as if others lack rights they in fact possess. But I doubt we can plausibly flesh out the *Ur-Idea* in terms of either of these deliberative failings. Consider, first, the evidence-relative failure: X fails to give Y's right the weight in her deliberations that she ought given her evidence. The problem is that the mistaken aggressors in *Scare the Bear*, *Warehouse*, and *Waiting on Trolley* needn't exhibit this kind of failure. In each of those cases, we can simply stipulate in our description that the mistaken aggressor does not fail to give appropriate deliberative weight to the rights of their victims in light of their evidence. The same goes for the *Resident* case. Appeal to an evidential-deliberative failing cannot, then, explain why the mistakes in these various cases renders the aggressor liable to defensive force.

Consider the fact-relative, deliberative failure: X fails to give Y's right the weight in her deliberations that she ought to given the facts. Appeal to this sort of failure fares a bit better. In *Resident*, the mistaken aggressor does engage in a deliberative failing in this sense. She fails to give proper weight in her deliberations to the rights the innocent twin actually has. If the resident were aware of all the facts, we would clearly blame her for the way she deliberates. The same goes for the mistaken aggressor in the *Waiting on Trolley* cases: since there is, in fact, no one on the primary tracks, and since Brianna is not in fact liable to be harmed, Alice fails to treat her rights with their due weight when she deliberates as she does. The problem, however, is that there is no such failure in a case like *Warehouse*. In *Warehouse*, Adam does not fail to give Beatrice and Charlie's rights the weight he ought to. He correctly judges that they have full rights against being harmed. His failing is not that he fails to properly appreciate their rights in his deliberations, but consists only in his mistaken belief that those rights are overridden by a greater benefit.²⁹

²⁹It's less clear how to characterize the failure in *Scare the Bear*. In one sense, Hiker does

There is, however, another available interpretation of the as-if phenomenon — distinct both from Quong’s interpretation and the deliberative interpretations. Call it the *objective* interpretation: to treat someone as if they lack a certain right is simply to act in a way that in fact threatens that right without a fact-relative justification for doing so.

This interpretation is faithful to the ‘as if’ language in the following way. When we say that A acts ‘as if B lacks a right,’ we’re often saying something about *how A’s action treats B*, rather than something about A’s mental states or deliberative process. Specifically, we are saying that A’s action infringes B’s rights without a fact-relative justification for doing so. Why the justification qualification? Because a fact-relative justification transforms a rights violation into a permissible infringement, and it seems implausible to describe a permissible infringement of a right as a way of acting as if that right isn’t in force. On the contrary, when A permissibly infringes B’s rights, she is acting in a way consistent with the force of B’s rights. It is only when there is no such justification that transgressing Bob’s rights treats him in a way that is inconsistent with the force of those rights.

The objective understanding of the as-if phenomenon has a decisive advantage over Quong’s evidential understanding. If we flesh out the *Ur-Idea* in these objective terms, rather than in evidential terms, a theory of liability emerges that accommodates the full range of liability-grounding mistakes canvassed in this paper. Resident, Scare the Bear, Warehouse, Waiting on Trolley — in each of these cases the agent mistakenly treats someone as if they lack rights in the objective sense.

But notice what it would mean to endorse the *Ur-Idea* so understood. It *not* fail to give Alice and Bob’s rights the deliberative weight he ought to in light of the facts: it isn’t part of his deliberations that Alice and Bob have less weighty rights than they in fact do. But there is a sense in which Hiker does fail to give their rights the deliberative weight he ought to: if he knew the facts, Alice and Bob’s rights should figure more prominently in his deliberations. This is thus one sense in which Hiker fails to take his victim’s rights into his deliberations that isn’t present in the Warehouse case, where Adam’s victims figure as prominently as they ought to in his deliberations.

would be to come full circle: it would be to endorse nothing other than Thomson's classic VIOLATION theory. VIOLATION, recall, just is the claim that a person is liable to defensive harm when they pose a threat to someone's rights. This is just another way of saying that a person is liable when they act (in the objective sense) as if someone lacks a right they in fact possess.³⁰

Quong says:

What matters is whether a person, *judged from the evidence-relative standpoint*, acts as if others lack rights against harm. When, and only when, our actions have this particular feature — treating others with something less than the concern and respect they are due — we make ourselves liable to defensive harm.³¹

VIOLATION just is the view expressed by this quote, replacing 'evidence-relative standpoint' with 'fact-relative standpoint.' MORAL STATUS and VIOLATION can both be understood as sharing the same underlying idea that liability follows from a failure to treat others with "the concern and respect they are due." They merely differ in their choice between an evidence-relative and fact-relative form of this failure.

I'm not claiming that we should endorse VIOLATION. My claim is just that in as much as one finds Quong's *Ur-Idea* intuitively compelling, one arguably has more reason to embrace Thomson's VIOLATION account on that basis than Quong's MORAL STATUS account.

³⁰To get RESPONSIBILITY or CULPABILITY we would need to soften the *Ur-Idea*. We would need to say that acting (in the objective sense) as if someone lacks a right they in fact possess is *not sufficient* for liability. We would need to add a responsibility or culpability condition.

³¹Quong, *Defensive Force*: 41.