Liability Without Fault

Ketan Ramakrishnan

Balliol College
University of Oxford



Submitted for the degree of
Doctor of Philosophy
Michaelmas 2020

PREFATORY NOTE AND GROVELING APOLOGY TO READERS

This thesis investigates the question of when, if ever, an agent who threatens to harm another person becomes liable to defensive harm although he is acting without any sort of moral, cognitive, or epistemic fault. In other words, it investigates *strict liability to defensive harm.* The thesis consists of two articles.[1]

The first article argues that there are close extensional connections between fact-relative duties against harm, evidence-relative duties against risk imposition, evidence-relative obligations to internalize risk, and strict defensive liability. It exploits these connections in order to argue against some prominent theories of strict defensive liability in the current philosophical literature, and to intervene in some important debates about its scope. And it tentatively proposes the outline of an alternative theory that might explain these extensional connections. According to this theory, strict defensive liability and obligations to internalize risk are both grounded in the stringency of the evidence-relative duties against risk imposition which a threatener is infringing. These duties are, in turn, presumptively grounded in the stringency of the underlying fact-relative duty against harm whose infringement is at stake. But all sorts of considerations – such as the relevant victim's consent to the relevant risk of harm, or the fact that she is imposing or has imposed a comparable risk of harm on the threatener, or the fact that the threatener is acting for the benefit of other people, or the fact that the victim could easily have avoided suffering the harm – might sever this presumptive connection between fact-relative duties against harm and evidence-relative duties against risk imposition. Much of the philosophical action lies in identifying the considerations that might play such a role and explaining how, exactly, they do so.

The second article examines the strict defensive liability of agents who are imposing only astronomically small or freakishly unusual risks of harm on their victims. It is often assumed that such agents are liable to defensive harm – if they are liable at all – only because their bodies are moving in harmful

[1] I don't know if there is any feasible way of getting the ridiculously dense and intricate material in this thesis published. But my current plan of attack is to submit it in the form of two large articles, and see if any philosophy journal is willing to bit. (The various parts of the thesis are too closely related, I suspect, to divide it up into several smaller articles. And the arguments are perhaps just too tentative for publication in the form of a monograph, which is in any event difficult for junior scholars to secure.) I have used this format in this thesis, as well, in part because I would be interested to hear your thoughts on whether the format is unwieldy, difficult to follow, or otherwise undesirable.

ways, rather than because of their responsible actions. Reflection on obligations to internalize risk suggests, instead, that at least some of these agents become liable to severe defensive harm through the voluntary exercise of their responsible agency.

A much less talented philosopher once wrote, of his much less impressive work, "I hate this book."[2] I hate this thesis. It has taken me far too long to complete, and I am sorry for the burdens that this ridiculous delay has imposed on you, my readers. There are many important things that I have simply been unable to figure out. Some of this perplexity may derive from the difficulty and complexity of the subject matter, but to a large extent it is doubtless the result of my own philosophical limitations. And as I suggest at one point in the first article, I worry that the sum total of this work is little more than "[a]n expression of difficulties, uncertainties and doubts, arriving at no very definite conclusion".[3] I might have been able to resolve some of these difficulties if I had sat with the material for another couple of years, no doubt devising even more fantastical cases involving deadly gas releases and risk-internalizing magical buttons. But at some point the baby must be put to bed.

[2] V. Tadros, *Wrongs and Crimes* (Oxford: Oxford University Press, 2016): v.
[3] W. Prosser, "*Palsgraf* Revisited," *Michigan Law Review 5*2 (1953) 1: 1 - 33, 2.

TABLE OF CONTENTS

Strict Liability to Defensive Harm I: Internalizing Risk and Justifying Defense……1

Strict Liability to Defensive Harm II: The Specter of Absolute Liability…….93

STRICT LIABILITY TO DEFENSIVE HARM I: INTERNALIZING RISK AND JUSTIFYING DEFENSE

Ordinary moral thought recognizes a significant moral asymmetry between doing and allowing harm. So it seems impermissible, for example, for an impartial observer to kill one innocent stranger in order to save the life of another. But it is widely agreed that culpable threateners render themselves *liable* to harm: if it is necessary to inflict harm on a culpable threatener in order to prevent him from inflicting comparable harm on his innocent victim, the threatener will lose his ordinary moral rights against suffering the harm in question. And so, for example, even a wholly impartial observer may defensively kill a culpable threatener in order to save the life of his innocent victim.

There is much less agreement about the defensive liability of faultless threateners, agents who pose wrongful threats of harm to other innocent people but cannot be accused of any moral, epistemic, or cognitive failing in doing so.[1] Some philosophers believe that, since a faultless threatener is just as innocent as his victim, the moral claims of both parties remain wholly intact.[2] Thus it is impermissible, on their view, for an impartial observer to kill a faultless threatener in order to save the life of his victim: here, as in general, the duty not to kill one innocent stranger defeats the duty to save another.[3] To be sure, these philosophers generally allow, the *victim* of a faultless threatener is permitted to defend herself against his threat. But she enjoys this special defensive permission because of her special stake in the matter, the fact that she is the

[1] Little has been written about the defensive liability of threateners acting in a morally blameless but epistemically faulty manner. This is a difficult issue that I will not address in this article, except briefly at p. 29 below. Tort theorists have paid more attention to similar issues regarding corrective liability. See, for example, A. Ripstein, *Equality, Responsibility, and the Law* (Cambridge: Cambridge University Press, 1998): 84 - 87.

[2] See e.g. L. Alexander, "Recipe for a Theory of Self-Defense: The Ingredients, and Some Cooking Suggestions," in *The Ethics of Self-Defense,* eds. C. Coons and M. Weber (Oxford: Oxford University Press, 2014): 20 -50; K. Ferzan, "Culpable Aggression: The Basis of Moral Liability to Defensive Killing," *Ohio State Journal of Criminal Law* 9 (2012) 2: 669 - 697; D. Rodin, *War and Self-Defense* (Oxford: Oxford University Press, 2009). See also, more ambiguously, S. Burri, "Morally Permissible Risk Imposition and Liability to Defensive Harm," *Law and Philosophy* 39 (2020) 4: 381 - 408; S. Lazar, "Responsibility, Risk, and Killing in Self-Defense," *Ethics* 119 (2009) 4: 699 - 728.

[3] Like most other contemporary writers on the morality of defensive harm, I will make use of the distinction between *pro tanto* moral requirements (duties, obligations) and all-things-considered moral requirements (duties, obligations). In ordinary moral discourse, the terms "duty" and "obligation" are ambiguous; each can be used to denote either a *pro tanto* moral requirement or an all-things-considered moral requirement. But here I will adopt, for the sake of expository clarity, a uniform terminological convention whereby "duty" will refer exclusively to a *pro tanto* moral requirement and "obligation" will refer exclusively to an all-things-considered moral requirement. Sometimes I will speak of "*pro tanto* duties" and "all-things-considered obligations" but often I will omit the qualifiers.

very person who is about to suffer harm (or whose rights against harm are going to be infringed).[4] Impartial observers have no such special permission, and so they should simply refrain from intervention.

Other philosophers reject this suggestion of moral parity between faultless threatener and victim. On their view, when a faultless threatener reasonably chooses to risk harm to another innocent person, this choice is (or can be) a morally grave exercise of his responsible agency. If this risk of harm should threaten to materialize, therefore, the threatener in question can become liable to comparable defensive harm.[5] It is common ground among these philosophers that neither moral blameworthiness nor any sort of cognitive or epistemic fault is a precondition of defensive liability. But there are significant disagreements within this broad philosophical camp about exactly which faultless threateners incur defensive liability, and about the underlying moral principles that explain why such liability arises.

In short, there is considerable controversy about the existence, scope, and grounds of *strict liability* to defensive harm.[6] The debate about strict defensive liability has important practical implications, concerning issues such as the initiation and prosecution of war and the availability of compensatory damages in the law of torts. It also has broad relevance for the morality of defensive action, since the moral norms

---

[4] The harm-based understanding of this agent-relative prerogative is proposed in, for example, J. Quong, "Killing in Self-Defense," *Ethics* 119 (2009) 3: 507 - 537. The rights-based understanding of the prerogative is proposed in, for example, S. Bazargan, "Moral Coercion," *Philosophers' Imprint* 14 (2014) 11: 1 - 18, 11.

[5] There is a large literature defending views of this general kind. See, for example, S. Bazargan, "Killing Minimally Responsible Threats," *Ethics* 125 (2014) 1: 114 - 136; K. Draper, *War and Individual Rights* (Oxford: Oxford University Press, 2015); H. Frowe, *Defensive Killing* (Oxford: Oxford University Press, 2014); K. Gordon-Solomon, "What Makes a Person Morally Liable to Defensive Harm?" *Philosophy and Phenomenological Research* 97 (2018) 3: 543 - 567; J. McMahan, "The Basis of Moral Liability to Defensive Killing," *Philosophical Issues* 15 (2005): 386 - 405; J. McMahan, *Killing in War* (Oxford: Oxford University Press, 2011); J. McMahan, "Self-Defense Against Justified Threateners," in *How We Fight* eds. H. Frowe and G. Lang (Oxford: Oxford University Press, 2014): 104 – 137; M. Otsuka, "Killing the Innocent in Self-Defense," *Philosophy & Public Affairs* 23 (1994) 1: 74 - 94, p. 91; M. Otuska, "The Moral Responsibility Account of Liability to Defensive Killing," in *The Ethics of Self-Defense,* eds. C. Coons and M. Weber (Oxford: Oxford University Press, 2014): 51 – 69; J. Quong, *The Morality of Defensive Force* (Oxford: Oxford University Press, 2020); V. Tadros, *The Ends of Harm* (Oxford: Oxford University Press, 2011); V. Tadros, *Wrongs and Crimes* (Oxford: Oxford University Press, 2016); V. Tadros, *To Do, To Die, To Reason Why: Individual Ethics in War* (Oxford: Oxford University Press, 2020); J.J. Thomson, "Self-Defense," *Philosophy & Public Affairs* 20 (1991) 4: 283 – 31; A. Walen, *The Mechanics of Claims and Permissible Killing in War* (Oxford: Oxford University Press, 2020).

[6] I will use the term "strict liability" to refer to any liability incurred by a wholly faultless threatener, that is, a threatener whose action is belief- and evidence-relative permissible, and who performs this action on the basis of fully adequate moral deliberation. The term "strict liability" is more familiar in law and legal theory than in philosophical writing on the morality of defense. But in certain important respects, I believe, it is preferable to the terminology that philosophers writing in this area typically employ. These philosophers typically distinguish between culpable threateners, "minimally responsible threateners," and "non-responsible threats." For reasons I will articulate soon below, I believe this terminology is distorting and ought to be abandoned.

that govern defense against faultless threateners are substantially continuous with the moral norms that govern defense against threateners who are culpable or otherwise at fault. But this debate might also merit the attention of philosophers without any particular interest in war, tort law, or other such applied ethical domains. It implicates some of the most difficult and interesting questions in foundational moral theory, concerning the moral norms that govern imposing and distributing risks of harm.

My central aim in this article is to examine the close connections between strict defensive liability and obligations to internalize risks of harm. Several decades ago, some important articles in tort theory examined similar connections between obligations to internalize risk and strict corrective liability.[7] Obligations to internalize risk have received much less attention from philosophers working on the morality of defense.[8] Whether a faultless threatener is obligated to internalize the risk of harm that she is imposing on her victim (or would be, if she could) turns out to be an excellent proxy for whether she is liable to defensive harm. Both obligations to internalize risk and strict defensive liability are closely connected, in turn, to the stringency of the evidence-relative duty against risk imposition which the relevant threatener is infringing.[9] And that, in turn, is closely connected to the stringency of the fact-relative duty against harm that she is threatening to infringe. Reflecting on these connections constitutes a powerful method for

[7] See R. Epstein, "A Theory of Strict Liability," *Journal of Legal Studies* 2 (1973) 1: 151 - 204; D. McCarthy, "Liability and Risk," *Philosophy & Public Affairs* 25 (1996) 3: 238 - 262. Such arguments are criticized in S. Perry, "Libertarianism, Entitlement, and Responsibility," *Philosophy & Public Affairs* 26 (1997) 4: 351 - 396.

[8] That said, the general issue of the relationship between defensive liability and a threatener's duties and obligations has been insightfully discussed in an important body of work by philosophers such as Adil Haque, Adam Hosein, and, especially, Victor Tadros. See Haque, *Law and Morality at War;* A. Hosein, "Are Justified Aggressors A Threat to the Rights Theory of Self-Defense?" in *How We Fight: Ethics in War* eds. H. Frowe and G. Lang (Oxford: Oxford University Press, 2014): 87 – 103; V. Tadros, "Duty and Liability," *Utilitas* 24 (2012) 2: 259 – 277; Tadros, *The Ends of Harm*; Tadros, *Wrongs and Crimes*. I discuss these philosophers' views at several points in this article, especially at pp. 26 - 29 and pp. 76 - 79.

[9] Throughout the article, I will make use of the standard philosophical distinction between fact-relative, evidence-relative, and belief-relative moral requirements. (Since all of the threateners I will discuss are acting in a cognitively and epistemically faultless manner, with the exception of the threatener discussed at p. 29, their evidence-relative duties dovetail with their belief-relative duties, and I will typically refer only to the former.) The terminology may originate in D. Parfit, *On What Matters* (Oxford: Oxford University Press, 2011), pp. 150 - 174, but the basic distinction is of much older provenance. See e.g. T. Hurka, *British Ethical Theorists From Sidgwick to Ewing* (Oxford: Oxford University Press, 2014): 78 - 86. For an exposition and defense of the distinction's significance in the context of defensive liability, see V. Tadros, *The Ends of Harm* (Oxford: Oxford University Press, 2011), pp. 217 - 240.

thinking about the existence and scope of strict defensive liability. Explaining these connections is, by the same token, an important constraint on any considered philosophical account of this moral domain.[10]

In the first section of this article, I will elaborate these thoughts by constructing a *risk internalization argument* for the existence of strict defensive liability. This case-based argument will suggest two abstract extensional principles: the *risk internalization principle* and the *presumptive parity principle*. I will conclude the section by tentatively proposing a theoretical account of strict defensive liability that might explain why these extensional principles hold, and thereby explain the intuitive judgments which these principles seek to capture. On this account, strict liability to defensive harm is grounded in (and its scope is fixed by) the stringency of the evidence-relative duty against risk imposition which a faultless threatener is infringing. When an agent owes another person a stringent duty against subjecting him to some risk of harm, she will be presumptively obligated to impose a certain amount of risk on herself rather than infringe this duty. The agent's strict liability to defensive harm is a kind of functional surrogate for this evidence-relative obligation: both are grounded in the stringency of the threatener's evidence-relative duty against imposing the relevant risk. In turn, the stringency of an agent's evidence-relative duty against risk imposition is presumptively grounded in the stringency of the underlying, fact-relative duty against harm which the agent is poised to infringe. Yet various factors - such as the victim's consent to the relevant risk of harm, or her reciprocal imposition of a comparable risk on the threatener - can sever this presumptive relationship between fact-relative duties against harm and evidence-relative duties against risk imposition. These considerations can undercut the presumptive stringency of a threatener's evidence-relative duties against risk imposition, and thereby undercut her presumptive liability to defensive harm.

[10] Obligations to internalize risk are also an important moral issue in their own right. Philosophers have not much discussed it, despite the abundance of excellent philosophical work on other topics in the morality of risk imposition (in particular, the impartial distribution of expected harms and benefits among many persons). See e.g. D. McCarthy, "Rights, Explanation, and Risks," *Ethics* 107 (1997) 2: 205 - 222; J. Frick, "Contractualism and Social Risk," *Philosophy & Public Affairs* 43 (2015) 3: 175 - 223; A. Mahtani, "The *Ex Ante* Pareto Principle," *The Journal of Philosophy* 114 (2017) 6: 303 – 323; P. Tomlin, "On Limited Aggregation," *Philosophy & Public Affairs* 45 (2017) 3: 232 - 260; J. Horton, "Aggregation, Risk, and Reductio," *Ethics* 130 (2020) 4: 514 - 529; A. Walen, "Risks and Weak Aggregation: Why Different Models of Risk Suit Different Types of Cases," *Ethics* 131 (2020) 1: 62 - 86.

Whether or not this particular theory of strict defensive liability is correct, any adequate theory must ultimately explain the kinds of extensional principles - regarding connections between fact-relative duties against harm, evidence-relative duties against risk imposition, evidence-relative obligations to internalize risks of harm, and strict defensive liability - that I will propose, along with the concrete intuitive judgments about risk internalization that these principles seek to capture. Thinking about risk internalization thus provides an illuminating framework for assessing existing theories and claims about strict defensive liability, as I will seek to demonstrate in the remainder of the article. The fertility of the framework will be illustrated by the breadth of the discussion, or so I hope. But readers need not traverse its entirety: while the remaining sections are mutually reinforcing, and sequenced in an ordered progression, many of the sections can be read on their own.

In section two, I will examine the most prominent account of strict defensive liability in the current philosophical literature: the *moral responsibility account* advocated by Jeff McMahan, Michael Otsuka, Kaila Draper, and many others. I will offer some novel case-based criticisms of this view. Pressing these criticisms will lead me to discuss a broader and more abstract issue about strict defensive liability: its relationship to distributive fairness. Some writers reject strict defensive liability on the basis that it seems unfair that entirely blameless agents, who have acted with all due care, should forfeit their rights against harm. In sharp contrast, the moral responsibility account maintains that strict defensive liability is *grounded* in considerations of distributive fairness, of a broadly luck egalitarian character. But thinking about the connections between obligations to internalize risk and strict defensive liability erodes the plausibility of any such claim. Instead, like the primary moral duties from which they derive, obligations to internalize risk and strict defensive liability prove to be deeply *inimical* to distributive fairness. So it is a mistake to explain the former moral phenomena in terms of the latter. For the same reason, however, it is equally an error to *reject* strict defensive liability for lack of fairness, unless we are prepared to dismantle the wider edifice of deontological moral thought.

In the course of developing these points, I will examine *Mistaken Attacker*, one of the most widely discussed cases in the philosophical literature on strict defensive liability.[11] In this sort of case, a faultless threatener reasonably but mistakenly believes that an entirely innocent and unthreatening person is trying to murder her, and she moves to kill him in (perceived) self-defense; now an impartial observer can save the unthreatening victim only by defensively killing the faultless threatener. I will argue that this kind of threatener is indeed liable to defensive killing. But that is not because it is any fairer for the threatener to die than her victim, as the moral responsibility account supposes. It is instead because the threatener's duties against risking harm to the victim, by shooting and killing him, are highly stringent.

In section three, I will argue against the *moral status account,* which may be the moral responsibility account's most prominent rival in the contemporary literature on strict defensive liability. I will also consider an important objection to my risk internalization arguments: namely, that they presuppose an overly stringent conception of our duties against doing harm. In the course of responding to this objection, I will discuss a novel sort of hypothetical case, *Manipulative Attacker,* in which an agent faultlessly threatens to harm another innocent person by non-consensually using his body. Philosophers writing on strict defensive liability have not discussed such cases at any length. Instead they have concentrated on faultless threateners who are poised to inflict side-effect harm (as in Conscientious Driver) or eliminative harm (as in Mistaken Attacker).[12] I will argue that, even if some nonconsequentialists can coherently reject the existence of strict defensive liability in such cases, a wholesale rejection of strict defensive liability is in serious tension with moral commitments about manipulative harming that virtually all nonconsequentialists embrace.

---

[11] Cases of this kind were first offered in K. Draper, "Fairness and Self-Defense," *Social Theory and Practice* 19 (1993): 73 - 92, p. 84; and Otsuka, "Killing the Innocent in Self-Defense," p. 91. A similar case is offered in McMahan, "The Basis of Moral Liability," p. 387. See also Quong, "Liability to Defensive Harm," p. 53. I offer a more fully described version of the case below at p. 38, which incorporates elements from these different versions.

[12] An agent *eliminatively* harms another person when she harms him as a means of neutralizing a threat that he poses or appears to pose. An agent *opportunistically* harms another person when she harms him in order to exploit some opportunity that his body provides. An agent *manipulatively* harms another person when she harmfully uses his body as a means of securing her ends. These distinctions trace back to W. Quinn, "Actions, Intentions and Consequences: The Doctrine of Double Effect," in *Morality and Action* (Cambridge: Cambridge University Press, 1993), pp. 175-194. Quinn did not, however, distinguish conceptually or morally between manipulative and opportunistic harming. On the importance of the distinction, see for example V. Tadros, "Wrongful Intentions Without Closeness," *Philosophy & Public Affairs* (2015): 52 - 74.

In section four, I will closely examine *Conscientious Driver,* the most controversial putative example of strict defensive liability. In this sort of case, an agent is responsibly driving her car around town, thereby imposing very small risks of harm on nearby pedestrians, for the sake of obtaining some modest personal benefit (such as seeing a film at the local cinema). Unfortunately, one of these small risks materializes: the car suffers a freak malfunction and spins out of control, and it will soon hit and kill a pedestrian unless an impartial observer blows it up, killing the conscientious driver inside. Drawing on the work of the first section, I will argue that, on certain reasonable specifications of the case's details, the conscientious driver is indeed liable to defensive killing. On other specifications of the case's details, however, the conscientious driver may *not* be liable. Cases of this general kind thus exhibit greater diversity in their essential moral dynamics than many philosophers have supposed. This diversity derives from the multiplicity of considerations - involving issues such as reciprocal risk imposition, and the state or community's normative powers in respect of public property - that may bear on the stringency of evidence-relative duties against risking harm through car driving.

My focus in the majority of the article is, like most of the philosophical literature on strict defensive liability, on faultless threateners who are acting entirely for their own benefit (and with a purely self-regarding justification). But in section five I will examine *strict liability for altruistic wrongs.* In other words, I will discuss the liability of faultless threateners who are acting purely or partly for the benefit of other people (and with a purely or partly other-regarding justification). Extending the article's analytical framework to strict defensive liability for altruistic wrongs will further support its plausibility. And it will provide the outlines of an approach to this issue that is, in certain important respects, more compelling than the philosophical approaches currently on offer. The extension is only tentative and incomplete, however, for it cannot comprehensively grapple with the many factors that plausibly bear on the stringency of an altruistic threatener's duties against risk imposition.

In section six, I will turn to the bearing of *avoidability* on defensive liability. Philosophers such as Thomas Scanlon, Niko Kolodny, and Victor Tadros have influentially proposed that a threatener's liability to defensive harm is substantially or even entirely determined by comparing his opportunity to avoid

suffering harm with the opportunity enjoyed by his victim. I will argue that avoidability has much less bearing on defensive liability (including strict defensive liability) than such views envisage. But articulating avoidability's relevance to defensive liability with any precision will turn out to be an extremely difficult task. Here as elsewhere, attention to the relationship between duties against risk imposition and defensive liability can fruitfully guide our inquiry without fully resolving it.

The work of this article develops a picture of strict defensive liability, and its relationship to the morality of inflicting harm and internalizing risk, that is plausible both in its abstract form and its concrete implications. In the companion to this article, I will turn to an aspect of this picture that is perhaps less welcome. Reflection on the morality of risk internalization generates strong pressure to adopt an extremely expansive view of strict defensive liability. My guiding examples will be familiar: I will discuss characters such as Judith Thomson's *light flipper,* who threatens to kill her neighbor by flipping the light switch in her home and triggering a freak burst of electricity; Jeff McMahan's *phone operator,* whose phone has been programmed by a terrorist such that it will detonate a bomb, killing another innocent person, when it is dialed; and Robert Nozick's *falling man,* whose body is hurled at a small woman trapped at the bottom of a well, such that his body will soon crush and kill her.[13] Philosophers have extensively debated whether impartial observers may defensively kill such characters. And it is typically assumed that, if defensively killing these characters is indeed permissible, that is simply because of the threats their bodies pose. This is assumption is sufficiently widespread that it has been encrusted into the standard philosophical label for such characters: they are typically classified as "non-responsible threats."

Against this assumption, I will develop a risk internalization argument for thinking that the responsible *actions* of these threateners render them liable to defensive harm.[14] More generally, once we closely examine the connections between obligations to internalize risk and liability to defensive harm, we come under significant reflective pressure to accept that strict defensive liability is *absolute,* in the following

[13] McMahan, "The Basis of Moral Liability," p. 397 and *Killing in War,* p. 165; R. Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1997): 34; J. Thomson, *The Realm of Rights* (Cambridge: Harvard University Press, 1990): 229.
[14] Others have suggested the possibility. See e.g. Alexander, "Recipe for a Theory of Self-Defense," p. 28; Ferzan, "Culpable Aggression," p. 682; Walen, *The Mechanics of Claims,* p. 157.

sense: anyone whose self-interested choice threatens to harm another innocent person can become liable to suffer significant defensive harm, no matter how astronomically improbable the harm he threatens may be. This conclusion cuts against the initial intuitive sensibilities of many philosophers, including myself. But it will prove difficult to resist on reflection. Partly in an attempt to do so, I will develop some views about strict defensive liability - *the reciprocal risk account,* the *inherent risk account,* and *the expected beneficiary account* - that may allow us to narrow the class of threateners who incur strict defensive liability for imposing only astronomically small risks of harm. But none of these views, independently or in tandem, will allow us to establish a categorical exemption from strict defensive liability for all such threateners. And the more we think about the morality of internalizing astronomically small risks, the less plausible such a categorical exemption becomes. Liability to severe defensive harm may be an inescapable risk of rational agency, which inheres in our actions no matter how innocuous they reasonably appear.

Before proceeding, I should briefly discuss a few important conceptual and methodological preliminaries. First, in what follows I will largely ignore what might be called the *fine structure* of liability to harm. As I have said, it is morally impermissible for an impartial observer to inflict harm on a faultless threatener, in order to save his innocent victim from equivalent harm, unless the threatener's ordinary moral protections against being harmed have been compromised in some significant way. Along with most other philosophers writing on defensive harm, I will assume that the relevant change consists in the complete loss or forfeiture of the threatener's rights against suffering the relevant defensive harm.[15] It is possible, however, that the necessary moral change will sometimes or always take a different form.[16] My concern here is to determine when and why such a change occurs, not to elucidate its precise structure. Thus I will simply

[15] See e.g. McMahan, "The Basis of Moral Liability to Defensive Killing"; Quong, *The Morality of Defensive Force;* J. Thomson, "Self-Defense," *Philosophy & Public Affairs* 20 (1991) 4: 283 - 310.

[16] To illustrate, suppose we are trying to explain why an impartial observer may defensively kill a certain threatener, in order to save the life of the threatener's innocent victim. We might do so by positing that the threatener has lost his right against being killed (or, what is much the same, that he lacks such a right in these circumstances). But we might instead hold that the threatener retains his right against killed, and loses only his right against suffering some less-than-lethal but still significant harm - such that it is permissible, all things considered, to defensively kill him. Alternatively, we might hold that the threatener's ordinary rights against being killed drastically *weaken,* enough so that even impartial observers may permissibly infringe these rights in order to save his victim. Or perhaps we might posit some complex combination of rights loss and rights weakening.

adopt the common assumption that a threatener's liability to suffer a given defensive harm is best understood as the complete loss of his rights against suffering that harm (as a defensive measure, in the circumstances at hand). I believe my arguments are compatible, however, with other plausible views about the fine structure of defensive liability.

Second, I will mostly ignore self-defense (as opposed to other-defense) against faultless threateners. As I have mentioned, the prospective victim of a threatener may sometimes enjoy a special permission to defensively harm him, in order to save herself from suffering harm, even if the threatener's ordinary moral rights against being harmed remain entirely intact. By contrast, impartial observers do not possess any such special permission; they may harm a faultless threatener only if his ordinary moral rights against being harmed have been compromised in some way. By focusing on the defensive permissions of impartial observers rather than victims, therefore, we can more perspicuously isolate the conditions of liability to defensive harm.

Third, throughout the article, I will rely on an important philosophical distinction between the *stringency* of an agent's fact-relative duty against harm (i.e. the amount of harm she is presumptively fact-relative obligated to allow herself to suffer, or impose on herself, in order to refrain from infringing this duty)[17] and the *strength* of an agent's fact-relative duty (i.e. how much weight this duty has when it conflicts with countervailing moral considerations, such as moral duties owed to other people or moral reasons arising from impersonal moral values). The distinction between stringency and strength has been recognized by several philosophers, using various terminological schemes.[18] But these different species of a duty's weight are sometimes lumped together.[19] In my view it is important to keep them apart, for they can diverge in important ways. It is also important to note that the stringency of duties against harm will often be *context-*

[17] These amounts of harm might sometimes diverge. We might accordingly distinguish between the *passive stringency* of a duty and the duty's *active stringency.* While I will not explicitly draw this distinction in what follows, I will be principally concerned with active stringency.

[18] F.M. Kamm was the first to observe the basic distinction, in "Supererogation and Obligation," *The Journal of Philosophy* 82 (1985) 3: 118 - 138. See also S. Lazar, "Associative Duties and the Ethics of Killing in War," *Journal of Practical Ethics* 1 (2013) 1: 3 - 48; D. Parfit, *On What Matters* vol. 3 (Oxford: Oxford University Press, 2017): 369 - 371.

[19] See e.g. Quong, *The Morality of Defensive Force,* esp. pp. 107 - 114.

sensitive.[20] Put another away, how much harm an agent must impose on herself, in order to refrain from infringing some duty against harm that she owes to another person, can depend on various features of her situation apart from the magnitude and character of the relevant harm. It may depend, for example, on whether the agent's justification for acting is self-regarding or other-regarding in character. We will observe this context-sensitivity of fact-relative duties against harm - and, similarly, the context-sensitivity of evidence-relative duties against harm - at several important junctures in this article.

As compared to the stringency of fact-relative duties against harm, the stringency of evidence-relative duties against imposing *risks* of harm has been barely discussed. The stringency of an evidence-relative duty is a matter of how much risk an agent must accept (or impose on himself) in order to refrain from infringing this duty. I will try to show that we need some notion of this kind, and that it plays an important role in explaining strict defensive liability. In particular, it is the stringency of a threatening agent's evidence-relative duty against imposing a given risk of harm that determines whether she is obligated to internalize this risk (if she can), and whether she is liable to a defensive measure that will, in effect, force her to internalize her this risk. Or so I will soon argue.

I. Risk Internalization and Defensive Liability

In this section I will offer a risk internalization argument for the existence of strict defensive liability. I will then extract from this argument some general extensional principles about when, exactly, strict defensive liability arises. Finally, I will offer an account that might explain such extensional principles, along with the intuitive judgments about risk internalization and defensive liability which these principles capture.

We begin with a case in which it is morally impermissible for an agent to harm another person in order to save himself from equivalent harm. Consider

> *Sickly Driver 1:* Threatener can save himself from dying only by immediately driving himself to the hospital. As he knows, however, his car will unavoidably suffer a mechanical failure and spin out of control along the way. It will thus hit and kill a pedestrian, Victim, who is walking on a footpath by

[20] See generally Lazar, "Associative Duties"; Kamm, "Supererogation and Obligation."

the side of the road. After this incident, Threatener knows, he will be able to walk the rest of the distance to the hospital and get there in time for the necessary medical treatment.

It seems morally impermissible, in both the fact-relative and evidence-relative senses of the term, for Threatener to drive himself to the hospital in this case. His *pro tanto* duty against hitting and killing Victim appears to be highly stringent. It is sufficiently stringent that he must not infringe this duty even in order to save his own life. Of course, if Threatener *does* decide to infringe this duty, his decision may be somewhat understandable, given the dire circumstances and considerable psychological pressure under which he is acting. But these facts do not seem to render his action morally permissible, even if they partly excuse it.[21]

We next consider a variation of this initial case, in which the threatener imposes a comparatively small risk of harm on his victim:

> *Sickly Driver 2:* Threatener will almost certainly die unless he drives himself to the hospital for treatment (in which case he will almost certainly live). Unfortunately, as Threatener knows, there is a 1% chance that his car will spin out of control along the way, thereby hitting and killing a pedestrian, Victim, who is walking on a footpath by the side of the road. Threatener decides to drive to the hospital anyway, in order to save his life.

Threatener's action here seems evidence-relative permissible. Of course, if Threatener will in fact kill Victim by driving to the hospital, then doing so is a fact-relative impermissible violation of Victim's right against being killed. But it is hard to deny that Threatener may permissibly (in the evidence-relative sense) impose *some* risk of death on Victim, in order to spare himself from a near certainty of death. And here the risk is sufficiently small that Threatener is plausibly permitted to impose it.

As we next observe, however, Threatener may impose this risk on Victim *only if he cannot internalize the risk*, by imposing it on himself. If he can internalize this risk, he is obligated to do so:

---

[21] Some philosophers would reject this judgment. For these philosophers believe that each of us has a general, agent-relative prerogative to harm another person (so long as this does not use his body or property as a means), in order to avoid equivalent personal harm. Like the majority of philosophers writing on defensive liability, I do not believe that any such general prerogative exists. But risk internalization arguments for strict defensive liability can be devised even for philosophers who endorse such a prerogative, as I will explain in section three. See below, pp. 43 – 51.

Sickly Driver 3: As in Sickly Driver 2. However, before driving to the hospital, Threatener can install a device on his car that will continuously monitor its immediate surroundings and blow the car up if (and only if) it is about to hit and kill a pedestrian. Installing this risk-internalizing device will virtually eliminate the 1% chance that Threatener will kill Victim by driving to the hospital, and generate a precisely equivalent chance that Threatener will kill *himself* by driving to the hospital. Threatener has no other means of internalizing the risk he is imposing on Victim.

It seems highly plausible that Threatener is morally obligated (in the evidence-relative sense) to install this risk-internalizing mechanical device on his car before he drives to the hospital.[22] Intuitively, it is not permissible for Threatener to impose this risk of death on Victim, for his own expected benefit, when he can acquire the same expected benefit by imposing an equivalent risk of death on himself.[23] To be sure, Threatener would not be obligated to internalize this risk at *any* cost to himself. If installing the device were an extremely dangerous procedure - one that would likely cost him several limbs - it would be permissible for Threatener to refrain from installing the device. But in the case at hand, Threatener can install the device, and thereby internalize the risk of death he is imposing on Victim, without incurring any such further cost. It seems clear to me that he is morally obligated to do so.[24]

This judgment gains further support from reflecting on other cases in the vicinity. Suppose that a Threatener can save himself from a fate much worse than death either by driving into a tree, killing himself, or by driving towards a Victim who is walking by the side of the road, hitting and killing her (*Choice of Lethal*

---

[22] But this judgment, and the related judgments in the next several paragraphs, are inconsistent with recognizing a general, agent-relative prerogative to do harm. See above, footnote 21, and section three below, pp. 43 – 51.

[23] Compare the remarks on strict corrective liability in McCarthy, "Liability and Risk," esp. pp. 242, 244; Epstein, "A Theory of Strict Liability," esp. pp. 159 - 160.

[24] This judgment is consistent with Anti Kauppinen's view that (certain) self-driving cars should be programmed so as to impose, on their users, the risks of harm that using such cars entails. See "Who Should Bear the Risk When Self-Driving Cars Crash?" *Journal of Applied Philosophy* (forthcoming). But Kauppinen defends this conclusion on the basis of arguments - that using such a self-driving car is *reasonably avoidable*, because it is not unduly costly to refrain from doing so, and that it is more *fair,* in a luck egalitarian sense, for the user to suffer harm than his victim - that are different from my own, and which I would reject. One obvious divergence between our views arises in the present context: in the Sickly Driver cases, the faultless threatener involved cannot reasonably avoid setting his car on the road. So Kauppinen's view would seem to exempt this driver from the obligation to internalize the risk of death that his action is imposing on his victim. On my reasons for resisting the appeal to luck egalitarian fairness in this context, see below, pp. 36 - 43. On avoidability's limited bearing on liability, see below, pp. 80 – 92.

Harms). It seems clear that Threatener is obligated to choose the former option: he may not foist the terrible cost of saving himself from a fate worse than death onto another innocent person, when he can instead procure this personal benefit by imposing an equivalent cost on himself. This seems true even if the fate that Threatener is trying to escape is sufficiently terrible that, if he had no other way to escape it, he would be permitted to hit and kill Victim (as a side-effect) in order to do so.[25]

The preceding case involves a fact-relative obligation to internalize certain harm, but the same analysis applies to evidence-relative obligations to internalize *risks* of harm. For suppose that Threatener can save himself from a near certainty of a fate worse than death either by driving towards (and possibly into) a tree, imposing a 1% risk of death on himself, or by driving towards (and possibly into) a Victim, imposing a 1% risk of death on her (*Choice of Lethal Risks*). It seems clear that he must choose the former option: he may not foist the terrible (expected) cost of his self-benefiting action onto this other innocent person when he can instead bear the cost himself. Matters might be different if, for example, Victim had consented to the imposition of this 1% risk, or if she had previously imposed some comparable risk of harm upon Threatener. But absent any such undercutting factor, it seems clear to me that Threatener is evidence-relative obligated to internalize any given risk of killing Victim that he is creating in order to save himself from a fate worse than death.

If Threatener is indeed obligated to internalize the 1% risk of death in this last case, Choice of Lethal Risks, then it is difficult to deny that Threatener is obligated to internalize the 1% risk of death in Sickly Driver 3. The only relevant difference between these cases is that, whereas the former threatener is imposing a 1% risk of death on another person in order to save himself from a fate *worse* than death, the latter threatener is imposing a 1% risk of death on another person in order to save himself from death. It is implausible that this difference should render the former threatener's moral position less favorable than the

[25] I assume that *some* fate worse than death is sufficiently terrible that it would be permissible to infringe another innocent person's right against being killed (as a side-effect) in order to avoid it. Depending on one's intuitive sensibilities, imagining such a fate might require entering the realm of science fiction.

latter's. The more plausible view is that each threatener is obligated to internalize the (expected) cost of his self-benefiting action.

Reflection on these further cases thus buttresses the intuition that Threatener is obligated to internalize the risk of death in Sickly Driver 3. Together, moreover, these various intuitive judgments suggest a plausible general principle:

> THE PRESUMPTIVE PARITY PRINCIPLE: Suppose that an agent A's fact-relative duty against inflicting some harm on another person P is sufficiently stringent that A would be fact-relative obligated to impose this harm on himself (when acting entirely for his own benefit, with an entirely self-regarding fact-relative justification) rather than impose it on P. Then, absent some undercutting moral consideration, A's evidence-relative duty against imposing any given *risk* of this harm on P is sufficiently stringent that he would be evidence-relative obligated to impose this risk on himself (when acting entirely for his own benefit, with an entirely self-regarding evidence-relative justification), rather than impose it on P.

The basic idea behind this principle is simple: if an agent has a stringent fact-relative duty against inflicting some harm, then he has - presumptively - a comparably stringent evidence-relative duty against imposing any given *risk* of this harm. And so, if an agent would be fact-relative obligated to internalize a certain harm he will presumptively be evidence-relative obligated to internalize any given risk of this harm. [26]

This extensional principle captures the various judgments we have discussed. Just as Threatener is fact-relative obligated to internalize the lethal harm in Choice of Lethal Harms, so too Threatener is evidence-relative obligated to internalize the small risk of lethal harm in Choice of Lethal Risks. So too, Threatener is evidence-relative obligated to internalize the small risk of lethal harm in Sickly Driver 3. All of these cases involve the same kind of fact-relative duty against harm. And in all of these cases the threatener is risking the infringement of this duty entirely for his own benefit (and with only a self-regarding evidence-

[26] This formulation is a simplification, because the presumptive parity principle does not address the duties of agents acting entirely or partly for the benefit of other people (and with partly or entirely other-regarding justification). As we will see later in this section, and more extensively below at pp. 73 - 80, it is more difficult to formulate a general account of such duties.

relative justification). In all of these cases, therefore, the threatener is obligated to internalize the relevant harm, or any given risk of it, if he can.

In the next step of the risk internalization argument, we suppose that a human third party, rather than a mechanical device, can be utilized as Threatener's risk-internalizing mechanism:

> *Sickly Driver 4:* As in Sickly Driver 3, except that Threatener does not have any mechanical risk-internalizing device available to him. However, he *can* call up an impartial Observer, who happens to be an extremely skilled military veteran, and ask that she defend against the risk of harm that Threatener is imposing on Victim. The procedure will work as follows: Observer will stand atop a nearby hill, monitor Threatener's car as he drives to the hospital, and blow the car up if (and only if) it is about to hit and kill Victim. By requesting this defensive course of action from Observer, Threatener will virtually eliminate the 1% chance that he will kill Victim by driving to the hospital, and simultaneously incur a 1% chance that *he* will be killed.

It seems that Threatener is obligated to ask that Observer monitor his car, and blow it up if necessary, in order to prevent him from hitting and killing Victim. There does not appear to be any morally significant difference between Threatener's moral situation in this case and his moral situation in Sickly Driver 3.[27] In both cases, Threatener is obligated to internalize the risk of death that he is imposing on Victim for his own expected benefit - whether by activating a risk-internalizing mechanical device or requesting a functionally equivalent defensive measure from a suitably situated impartial observer. In an important sense, that is, Observer in Sickly Driver 4 is just a functional substitute for the risk-internalizing mechanical device in Sickly Driver 3.[28]

---

[27] Put more precisely, Observer is a functional substitute for the risk-internalizing device so far as Threatener's moral position is concerned. There may of course be other morally significant differences between the cases. For example, perhaps Observer has an agent-relative prerogative not to kill other (innocent) people, and is thus permitted to decline the request that Threatener is obligated to make of her. No such prerogative is enjoyed, quite obviously, by a mechanical device.

[28] Perhaps it is worth noting that Observer may permissibly monitor Threatener's car, and blow it up if necessary in order to protect Victim, *even if Threatener declines to request that Observer do so*. For whether Observer may permissibly protect Victim from being killed by Threatener does not plausibly depend on whether Threatener chooses to honor his moral obligation to request that she engage in this protective action. Of course, if Threatener is indeed obligated to make this request of Victim, but he knowingly declines to do so, he will be culpably responsibly for this omission and so become a blameworthy threatener, not a faultless one.

In the fifth step of the argument, we imagine that Threatener is unable to *ask* that Observer help him internalize the relevant risk, but she can *make* him do so, by engaging in a suitable course of defensive action on her own initiative. Consider

> *Sickly Driver 5:* As in Sickly Driver 4, but Threatener cannot ask Observer to monitor his car and blow it up, if necessary, in order to kill Victim. This is because, although Threatener is aware of Observer's presence, he lacks the means to communicate with her. Threatener decides to drive the hospital anyway, in order to save his life, although he will thereby impose a 1% risk of death on Victim. Observer sees this situation unfolding and knows all the relevant facts about it. She can protect Victim by monitoring Threatener's car and blowing it up, if necessary, in order to prevent him from hitting and killing Victim.

It seems permissible for Observer monitor Threatener's car and blow it up if necessary to save Victim. After all, Threatener would be obligated to request *this very defensive course of action* if he could. Threatener would be obligated to internalize, in this way, the risk of death that he is imposing on Victim. It seems implausible that Threatener's inability to make this morally obligatory request should make it impermissible for Observer to act in accordance with its terms (as she would otherwise be permitted to do). The point may be reinforced by considering the following variant of the case:

> *Sickly Driver 5**: As in Sickly Driver 5. However, initially Threatener *is* able to communicate with Observer, by calling her with his cellular phone. But then a powerful gust of wind blows the phone out of Threatener's hand and into the sea. If this gust had not blown by, it would have been permissible for Observer to monitor Threatener and kill him in defense of Victim (if necessary). For either Threatener would have requested this defensive course of action, or else he would have wrongfully declined to do so (in which case he would have become a culpable threat to Victim).

It would be implausible to suppose that this gust of wind, by depriving Threatener of his ability to communicate with Observer, should make Observer's otherwise permissible defensive course of action an impermissible one. But this gust of wind simply dramatizes the salient feature of the previous case: Threatener's inability to request that Observer monitor and defend against his threat to Victim does not

change whether Observer may permissibly engage in this defensive course of action. Instead, it seems more plausible to say, Observer may permissibly monitor and defend against Threatener's threat to Victim regardless of whether Threatener can, or does, ask her to do so.[29]

Let us consider a final variant of the case. Suppose that Threatener cannot request that Observer defensively monitor him (and kill him if necessary), *not* because he lacks a means of communicating with her, but simply because he is unaware that she is on the scene (*Sickly Driver 5***). Here Threatener faces an epistemic obstacle, rather than a physical one, to making a request that he would otherwise be obligated to make. But this difference does not, it seems to me, change the basic moral structure of the situation. Nor am I aware of any account of strict defensive liability (or argument against its existence) according to which this difference would indeed be morally relevant. So I think we should conclude that, in all of these cases - Sickly Driver 5, Sickly Driver 5*, and Sickly Driver 5** - the faultless threatener involved may be defensively killed. If this is correct, then Threatener is liable to defensive harm at Observer's hands. For, to repeat, it is typically impermissible for an impartial observer to kill one person simply in order to save another. In Sickly Driver 5 and its variants, therefore, Threatener's ordinary moral rights against being killed by Observer have been lost or significantly weakened.

These conclusions might be challenged on the following basis. Suppose it is granted that Threatener would be obligated, if he could, to internalize the risk of death that he is imposing on Victim: he would be obligated to install the risk-internalizing mechanical device on his car (as in Sickly Driver 3) or request that Observer defensively monitor and kill him if necessary to save Victim (as in Sickly Driver 4). Still, it might be asked, why does this fact about Threatener's moral position give Observer any reason for defensive intervention? If Threatener actually *does* request this defensive course of action (as in Sickly Driver 4), Observer may thereby acquire a reason to intervene, for requests are often reason-giving.[30] But in Sickly Driver 5 and its variants, Threatener has not made any such request. If no other reason for Observer's

[29] This is compatible with thinking that, if Threatener actually makes such a request, Observer will have *additional* reason to engage in this defensive course of action. I discuss the point immediately below.
[30] D. Enoch, "Giving Practical Reasons," *Philosophers' Imprint* 4 (2011) 11: 1 - 22.

defensive intervention can be found, therefore, the conclusion that intervention is permissible in Sickly Driver 5 and its variants may come under strain.

In response I would make the following observation. As a general matter, it seems that we have some reason to enable other people to comply with their moral duties, even when doing so will prove costly and unwelcome to them.[31] Suppose, for example, that Theresa can save her ship from being destroyed by a terrible storm only by latching it onto the dock of her neighbor, Vincenzo, which will unavoidably destroy his dock.[32] Unfortunately, while Theresa *also* owns a nearby dock, it is currently inaccessible to her. It is permissible for Theresa to use Vincenzo's dock, if that is her only available means to save her ship, and her material interest in saving her ship is sufficiently large. But suppose an impartial observer can enable Theresa to access her own dock - and thereby save her ship by destroying her own dock instead of Vincenzo's. It seems plausible that the observer has some moral reason to do this. And this is true, it appears, although the observer has no independent reason to favor the destruction of one party's dock over the other's. I am not sure why, exactly, agents have reason to be concerned with other people's duties, in this way. But this seems to be a general moral phenomenon. And one plausible instance of this phenomenon is the fact that Observer has some moral reason to prevent Threatener from infringing his duties to Victim in cases like Sickly Driver 5.

Let me recap the second half of the risk internalization argument. If my remarks about Sickly Driver 3 and Sickly Driver 4 are sound, then Threatener is morally obligated (if he can) to internalize the risk of harm that he is imposing on Victim, whether by using a mechanical device or by communicating a functionally equivalent request to an appropriately situated third-party such as Observer. If my remarks about Sickly Driver 5 and its variants are sound, then Observer may act in accordance with the terms of

[31] A particularly strong version of this view, according to which everyone's moral reasons are equally reason-giving for everyone else, is defended in J. Gardner, "Complicity and Causality," *Criminal Law and Philosophy* 1 (2): 127 - 141. One need not accept that strong version of the view in order to embrace the more limited claim I am endorsing.
[32] Cf. *Vincent v. Lake Erie Transportation Co.* 124 N.W. 221 (Minn. 1910).

this request, e*ven if Threatener cannot or does not actually make it*, whether because of a practical or epistemic obstacle. So Threatener is liable to lethal defensive harm at Observer's hands.[33]

This risk internalization argument thus supports the existence of strict defensive liability. It also suggests a plausible general principle about when strict defensive liability arises:

> THE RISK INTERNALIZATION PRINCIPLE: A faultless threatener is liable to defensive harm, in order to prevent a risk of equivalent harm that he is imposing on another person from materializing, just in case his evidence-relative duty against imposing this risk is sufficiently stringent that he would be obligated to internalize this risk if he could (absent defeating moral considerations).[34]

This principle captures the extensional connections between (a) Threatener's obligation to internalize the risk of harm he is imposing on Victim, by means of a mechanical device, in Sickly Driver 3, (b) Threatener's obligation to internalize the risk of harm he is imposing on Victim, by requesting a suitable course of

---

[33] It may be worth noting an alternative and complementary route from the premise that Threatener is obligated to install the risk-internalizing device, in Sickly Driver 3, to the conclusion that Threatener is liable to defensive harm in Sickly Driver 5 and its variants. Consider a version of Sickly Driver 3 in which Threatener has access to the risk-internalizing mechanical device but does not know how to install it on his car. He can, however, ask impartial observer Observer, a skilled mechanic, to install the device on his car before he drives to the hospital. It seems clear that Threatener must ask Observer to do this. It would be strange to maintain that Threatener must install the device himself, in Sickly Driver 3, but that he may permissibly decline to request that Observer install the device in this variant. And it is plausible that Observer may install the device on Threatener's car even if Threatener cannot request that she do, whether because of a physical obstacle (he lacks the means to communicate with her) or an epistemic one (he is unaware of her existence or presence on the scene). To maintain that Observer may install the device on Threatener's car, whether or not he can or does request that she do so, is to maintain that Threatener is liable to a certain form of defensive harm at Observer's hands. For by installing the device on Threatener's car, Observer in effect ensures that Threatener will, if necessary, be defensively harmed in the future. If Observer may perform the defensive measure of installing the risk-internalizing device on Threatener's car in this case - even though Threatener cannot request, and is not requesting, that she do so - then it is hard to deny that Observer may engage in the functionally equivalent defensive course of action in Sickly Driver 5, 5*, and 5**.

[34] The bracketed proviso at the end of the principle is intended to handle such cases such as the following:

> *Destruction of Manhattan:* As in Sickly Driver 4. However, a whimsical villain credibly threatens to destroy Manhattan if Threatener internalizes the risk he is imposing on Victim, whether by using a mechanical device or communicating a request for defensive monitoring to Observer. The whimsical villain will not destroy Manhattan if Observer defensively monitors Threatener (and kills him if necessary to save Victim) on her own initiative.

In light of the villain's threat, Threatener is not all-things-considered obligated to internalize the risk he is imposing on Victim. He is not, therefore, obligated to request defensive action from Observer. (In fact, he is obligated to *refrain* from requesting such intervention.) Nevertheless, Threatener is liable to a defensive measure that will in effect force him to internalize this risk, and Observer may engage in such a defensive measure.

defensive action from Observer, in Sickly Driver 4, and (c) Threatener's liability to defensive harm at Observer's hands in Sickly Driver 5 and its variants. Threatener is obligated to internalize the risk of death he is imposing on Victim, and he is liable to a defensive measure that will in effect force him to internalize this risk.[35]

The risk internalization principle also strikes me as inherently appealing. Both obligations to internalize risk and liability to defensive harm are plausibly understood as forms of substantive moral responsibility for the consequences of risk imposition.[36] It seems reasonable to suppose that these burdensome varieties of substantive responsibility should be closely connected. And so it is plausible to suppose that, if an agent's duty against imposing some risk on another person is sufficiently stringent that he would be obligated to internalize this risk, if he could, the agent is also liable to defensive measures that will in effect force him to internalize this risk.[37]

By contrast, if an agent's choice to risk harm to another person is *not* sufficiently morally significant that he would be obligated to internalize this risk, it is difficult to see why a third party with no stake in the matter may effectively force him to do so. In the following section we will consider several cases that exhibit this structure. But here is an initial and especially forceful example:

> *Altruistic Injection:* An agent, Threatener, sees an unconscious person, Victim, lying on the ground. Threatener knows that Victim has contracted a terrible virus. And she has excellent reason to believe that he will soon die of it, unless she immediately injects him with a medical antidote in her possession. As she knows, however, there is a 5% risk that Victim is in fact genetically immune to

---

[35] The risk internalization principle only applies to cases in which, in order to prevent a threatener from inflicting some harm, he must be subjected to equivalent defensive harm. I will limit myself to discussing such cases in this article. The principle can be naturally extended to other cases, I believe, but the extension may look a bit unwieldy, e.g.:

> A faultless threatener is liable to defensive harm H1, in order to prevent a risk of harm H2 that he is imposing on his innocent victim from materializing, just in case his evidence-relative duty against imposing this risk of H2 on her is sufficiently stringent that he would be obligated to instead impose an equivalent risk of H1 on himself, if he could (absent defeating moral considerations).

[36] For this usage of "substantive responsibility," see T.M. Scanlon, *What We Owe to Each Other* (Cambridge: Belknap University Press, 1998): 248 - 249.

[37] This claim must be qualified in order to deal with cases in which a threatener would *not* be obligated to internalize a risk, even if he could, in virtue of some extraneous defeating factor. See footnote 33, above.

the virus - in which case injecting him with the antidote will kill him, for genetic immunity to the virus is precisely correlated with being deathly allergic to its antidote. Threatener has no special tie to Victim or any other special reason to aid him. She nevertheless decides to treat Victim with the antidote, entirely for his own benefit. She has stuck in the needle and is about to inject the antidote. But as Observer (an impartial third party) happens to know, Victim is in fact genetically immune to the virus and allergic to the antidote. Observer knows, therefore, that Threatener is about to kill Victim by injecting him with the antidote. But Observer is unable to communicate with Threatener. So Observer can save Victim only by defensively shooting and killing Threatener with a sniper rifle, before she injects him with the antidote.[38]

Threatener's evidence-relative duty against risking death to Victim, in the course of providing him with the relevant expected benefit, does not seem highly stringent: she would not be obligated to internalize this risk even if she could. Suppose, for example, that Threatener could activate a risk-internalizing mechanical device that would monitor the situation and defensively kill her if and only if the device reliably determined that her injection of Victim was in fact going to kill him. It seems to me that Threatener would be permitted to refrain from activating this device before injecting Victim. If that is correct, the risk internalization principle implies that Threatener is not liable to be defensively killed. This conclusion seems correct as well. If Threatener is not obligated to internalize the risk of death she is imposing on Victim, it is implausible that an observer with no stake in the matter may effectively force her to do so.[39]

As I have said, Threatener's duty against imposing a 1% risk of death on Victim seems to have only limited stringency in this case. Intuitively, this seems to be largely because Threatener is imposing this risk entirely for the benefit of another person (indeed, the very person on whom she is imposing the risk). Comprehensively explaining the stringency of an altruistic faultless threatener's duties is a difficult task, as we will later see.[40] For the most part, however, we can avoid these difficulties in this article. For our principal

[38] Similar cases are offered in Tadros, *To Do, To Die, To Reason Why,* p. 140; Draper, *War and Individual Rights,* p. 74.
[39] It is much less clear whether *Victim* would be permitted to defensively kill Threatener (if, for example, he should wake up right before she injects him, and find himself unable to communicate with her).
[40] See below, pp. 73 – 80.

concern will be with faultless threateners who are acting entirely for their own benefit (and with justifications for acting that are entirely self-regarding). As we have seen, in our discussion of the Sickly Driver cases, the risk internalization principle implies that at least some of these threateners are liable to be defensively killed.

Importantly, however, the risk internalization principle also implies that some entirely self-interested faultless threateners are *not* liable to be killed. Consider, for example,

> *Consensual Russian Roulette:* A villain credibly threatens to kill Antonia unless she plays Russian roulette on Billy, thereby imposing a 20% risk of death on him. The villain also credibly threatens to kill Billy unless he plays Russian roulette on Antonia, thereby imposing a 20% risk of death on her. After discussing their predicament, the two parties arrive at an agreement. Antonia will waive her claim right against being subjected to a 20% risk of death, through Russian roulette, by Billy; in exchange, Billy will waive his own claim right against being subjected to a 20% risk of death, through Russian roulette, by Antonia. Each party is preparing the pull the trigger of his or her gun at the other. As an impartial observer knows, Antonia will in fact trigger her gun's loaded chamber, thereby killing Billy, but Billy will not in fact trigger his gun's loaded chamber. The observer can save Billy only by defensively shooting and killing Antonia with a sniper rifle.

The fact-relative duty against shooting and killing another person is, presumptively, a highly stringent one. Presumptively, for example, an agent would be fact-relative obligated to save herself from a fate much worse than death by shooting and killing herself, rather than by shooting and killing another innocent person.[41] Under the presumptive parity principle, therefore, Antonia presumptively has a stringent evidence-relative duty against imposing any given risk of death on Billy by shooting at him; this duty is sufficiently stringent that she is presumptively obligated to internalize any such risk.[42] But in the case at hand, Billy has waived his claim right against being subjected to the relevant (20%) risk of death. Thus Antonia lacks the stringent

[41] The same structure holds in situations where the stakes are lower and more realistic. Suppose, for example, the a villain will kill an innocent agent unless she shoots herself in the foot, permanently crippling it, or shoots and permanently cripples the foot of a passing stranger. The agent must do the former.

[42] If, for example, the villain had exclusively threatened Antonia's life, and given her the option of saving it either by playing Russian roulette on herself or by playing Russian roulette on Billy, she would have been morally obligated to choose the former option.

evidence-relative duty against subjecting him to this risk that she would typically possess.[43] Thus Antonia would *not* be obligated to internalize the risk she is imposing on Billy, even if she could. And so, the risk internalization principle implies, Antonia is not liable to a defensive measure that will in effect force her to internalize this risk. Hence the impartial observer may not defensively kill Antonia.

The case nicely illustrates the joint operation of the presumptive parity principle and the risk internalization principle. First we ask: how stringent is the fact-relative duty against harming another person which the threatener is about to infringe? Is it sufficiently stringent that the threatener would be fact-relative obligated (when acting entirely for her own benefit, with entirely self-regarding fact-relative justification) to impose this harm on herself rather impose it on the other person? If so, the threatener will presumptively be evidence-relative obligated to internalize the particular *risk* of harm she is imposing on this other person (when acting entirely for her own benefit, with entirely self-regarding evidence-relative justification). But this presumption can be undercut by some further factor, such as the other person's consent to the imposition of the relevant risk. If no such undercutting factor exists, however, the threatener will be obligated to internalize this risk (if she can) and liable to harmful defensive measures that will in effect force her to internalize it.

I do not pretend to have defended this analytical framework conclusively, but I hope that I have conveyed its appeal. The risk internalization principle and the presumptive parity principle constitute plausible abductive generalizations from the case-based judgments I have discussed, and both principles are inherently attractive. In an indirect way, the rest of the paper will constitute an extended defense of these principles, by showing how they can fruitfully guide our thinking about a range of examples and theories about strict defensive liability. Along the way, I will explicitly address some further objections. Before proceeding, however, there is an objection that I should immediately address.

---

[43] I have stipulated that Antonia's consent is given in exchange for good and equal consideration, to indulge in some legal jargon. For it is especially plausible that rights can be successfully waived by such consent (as opposed to consent given gratuitously, or less than equal consideration). The presence of the consideration may not essential, however, or else it may be important partly because it evidences the (sufficiently) voluntary character of the consensual waiver involved. On related issues in the morality of promise and contract, see J. Lewinsohn, "Paid on Both Sides: Quid Pro Quo Exchange and the Doctrine of Consideration," *Yale Law Journal* 129 (2020): 690 - 772.

The risk internalization principle and the presumptive parity principle are purely extensional in character: these principles articulate plausible extensional connections between certain moral phenomena without specifying the relations of explanatory dependence that hold between them. It might be argued that this silence about explanatory structure is a problem: extensional claims without deeper explanation are little philosophical progress. Not so, I believe. For if these extensional connections are indeed plausible, then *any* adequate explanatory account of strict defensive liability must comport with them. These extensional connections are thus important critical standards against which existing theories of strict defensive liability can be assessed. And this critical work can largely be done without offering any competing theory of defensive liability in their stead.

It will nevertheless be useful, before proceeding to that critical work, to set out the general form of explanatory account which I favor. It is a familiar thought, in judicial and scholarly writing about the law of torts, that compensatory damages aim to put the victim of a tortious wrong in "the same position as he would have been in if he had not sustained the wrong for which he is now getting his compensation or reparation".[44] The idea is that an injurer's duty of compensation is grounded in the primary duty which he breached by injuring his victim; the compensatory duty serves as a surrogate for the primary duty which has been infringed.[45] Whether all corrective or compensatory duties should be understood on this model is unclear. But it is plausible that this model, properly formulated and suitably developed, will capture a centrally important class of these moral duties.

In my view, strict liability to defensive harm serves an analogous function. Strict defensive liability - or, perhaps more accurately, the harmful defensive action which it licenses - is a sort of functional surrogate for compliance with an agent's primary evidence-relative duty against risking harm to his victim, in circumstances where the agent cannot or need not comply with that duty. In Sickly Driver 5, for example,

---

[44] *Livingstone v. Rawyards Coal Co.* (1880) 5 App. Cas. 25, 855. Similar formulations are ubiquitous. See the references in A. Burrows, *Remedies for Tort and Breach of Contract,* 4th ed. (Oxford: Oxford University Press, 2019): 38 fn. 9.

[45] In the literature on tort theory, views of this general kind have been influentially defended by John Gardner, Arthur Ripstein, and Ernest Weinrib. See J. Gardner "What is Tort Law For? Part 1: The Place of Corrective Justice," *Law and Philosophy* 30 (2011) 1: 1 - 50; A. Ripstein, *Private Wrongs* (Cambridge: Harvard University Press, 2016): 233 - 265; E. Weinrib, *The Idea of Private Law* (Oxford: Oxford University Press, 2012). These views are insightfully compared and assessed in S. Steel, "Compensation and Continuity" (unpublished manuscript).

Threatener may permissibly infringe his evidence-relative duty against imposing a 1% risk of death on Victim (by hitting and killing her with his car). He need not comply with this duty, given that he must infringe it in order to spare himself from a near certainty of death. Nevertheless, he remains liable to a defensive measure that will in effect force him to internalize this risk. For his duty against imposing this risk on Victim is sufficiently stringent that he would be obligated to internalize it if he could. In this way, Threatener's liability to defensive killing is grounded in the stringency of the evidence-relative duty, towards Victim, that he is permissibly infringing. More generally, I believe, the scope of defensive liability is fixed by the stringency of the evidence-relative duty against risk imposition which a threatener is about to infringe. That is, in turn, presumptively grounded in the stringency of the fact-relative duty against harm which the threatener is about to infringe.

The risk internalization principle and the presumptive parity principle are extensional expressions of this basic explanatory structure, or so I would submit. But this claim invites an obvious objection. Why should our explanatory account accord any role to evidence-relative duties against risk imposition at all? Why not cut out the explanatory middleman, so to speak, and instead ground a threatener's strict defensive liability *directly* in the stringency of the fact-relative duty against harm that he is about to infringe? Building on Victor Tadros' important work, Adil Haque has articulated the basic structure of such a view with great clarity:

> "In general, the stringency of a moral duty should determine the moral consequences of its violation. Suppose that some moral duty is sufficiently important that I am morally required to suffer some harm rather than violate that duty. If I threaten to violate that duty, then you will not wrong me by inflicting comparable harm on me in order to prevent me from doing so. I have no moral right to be free from such harm since I am morally required to suffer comparable harm rather

than violate my duty. In contrast, you would wrong me by inflicting a greater harm on me than I am morally required to suffer rather than violate my duty."[46]

On a plausible interpretation of these remarks, they suggest the following view: an agent is liable to a given defensive harm just when, and because, he is threatening to infringe a fact-relative *pro tanto* duty of sufficient stringency that he would be fact-relative obligated to suffer this harm rather than infringe this duty.[47] If this *fact-relative duty account* of strict defensive liability is correct, then evidence-relative duties against risk imposition have no role to play in explaining strict defensive liability. To be sure, evidence-relative obligations to internalize risk might still constitute a useful *proxy* for strict defensive liability. For these obligations are grounded in the stringency of evidence-relative duties against risk imposition, which are in turn presumptively grounded in the stringency of fact-relative duties against harm.[48] But evidence-relative duties against risk imposition would not, in this scheme, have any role in *grounding* strict defensive liability: they would be an explanatory epiphenomenon.

This view is simple and elegant. But we have already encountered two cases - Consensual Russian Roulette and Altruistic Injection - that tell against it. In each case, the relevant threatener has a stringent fact-relative duty against harming his victim, and this duty presumptively grounds a stringent evidence-relative duty against risking such harm to his victim. But the stringency of the evidence-relative duty seems to be undercut, in each case, by morally relevant facts about the situation: the fact that the victim has consented to the risk in question (in the first case) or the fact that the threatener is imposing this risk for the

[46] Haque, *Law and Morality at War,* pp. 70 - 71. See also Tadros, "Duty and Liability"; *The Ends of Harm*; *Wrongs and Crimes.* As Haque observes, Tadros tends to focus on the relationship between a threatener's liability and his duties to compensate for harms he has already caused or to undo harmful processes he has already set in motion. And Tadros discusses (I would add) duties against harm rather than duties against risk imposition. It is not obvious, however, that Tadros means to embrace the view that defensive liability is especially connected to remedial duties (rather than primary duties) and fact-relative duties against harm (rather than evidence-relative duties against risk imposition). And Tadros has been exploring the relationship between duties and liability for some time, so his commitments about this matter may have evolved. Thus my own arguments here might be incompatible with or complementary to his commitments, depending on how those commitments are best understood.

[47] Haque never articulates his view in such explicit terms. But the explicit formulation I have provided comports, I believe, with the wide range of insightful case-based arguments in the book.

[48] This might be true because, for example, a faultless threatener's fact-relative duty against harm directly grounds his strict defensive liability and *also* directly grounds his evidence-relative duties against imposing risks of this harm (which in turn ground presumptive obligations to internalize the relevant risks). This explanatory structure would create tight correlations between obligations to internalize risk and strict defensive liability, without grounding one in the other.

victim's own benefit, not his own (in the second case). So in neither case, it seems, would the threatener be obligated to internalize the risk of harm he is imposing on his victim, even if he could. And in neither case, it seems, is the threatener is liable to defensive harm that will in effect force him to internalize this risk.

If we ground strict defensive liability in the stringency of evidence-relative duties against risk imposition, we can explain these judgments. By contrast, if we ground strict defensive liability directly in the stringency of fact-relative duties against harm, it seems we will need to pronounce each threatener liable to be defensively killed. For in each case, the relevant threatener is about to infringe a highly stringent sort of fact-relative duty against harmful bodily intrusion. And in neither case has any fact about the situation reduced the ordinary stringency of this fact-relative duty. In Consensual Russian Roulette, for example, the victim has consented to waive his evidence-relative right against being subjected to a 1% risk of death, but he has *not* consented to waive his fact-relative right against being killed. The underlying fact-relative right remains intact.[49] In Altruistic Injection, there is no fact about the situation that might reduce the ordinary stringency of the threatener's fact-relative duty against killing the victim; it is not the case, for example, that by injecting (and so killing) the victim she will *actually* benefit him. Rather, it is the stringency of the threatener's evidence-relative duty against *risking* death to the victim that is reduced; what is important is that imposing this risk on the victim (by injecting him) is in his benefit *ex ante*, by the lights of the threatener's evidence.

One might contest the details of these claims, depending on how exactly one understands the idea of a fact-relative duty and distinguishes it from the idea of an evidence-relative duty. While it is intuitively plausible to draw a basic distinction of this kind, making it out with precision may take some work. Much of this work remains to be done. But my basic point is simply that we must ground strict defensive liability in the stringency of a threatener's duty against imposing the relevant risk - rather than directly in the

[49] The persistence of the victim's underlying fact-relative right against harm (or, what is the same, the threatener's underlying fact-relative duty) plausibly explains why, for example, the threatener *would* be obligated to internalize any higher risk of infringing this fact-relative duty than the 1% risk to which the victim has consented. The threatener's fact-relative duty grounds evidence-relative duties against imposing these higher risks, and these duties have not been waived by the victim.

stringency of his duty against actually imposing the relevant harm - on pain of generating implausible attributions of strict defensive liability in several cases (including others we will discuss below). This basic point will hold, I suspect, whatever the best precise analysis of fact-relative duties and evidence-relative duties turns out to be.

But there is another, more troubling question about my position: *why* should strict defensive liability be grounded in (evidence-relative) duties against risk imposition rather than (fact-relative) duties against harm? Risking harm is morally significant largely because actually inflicting harm is morally significant.[50] So it is not at all clear why strict defensive liability should be grounded in duties against risk imposition, rather than the underlying duties against harm. Some explanation of this fact seems called for. But I confess that I have no satisfying explanation to offer. The most I can say, at the moment, is this. Moral deliberation is always a matter of responding to risks, rather than certainties, of harm. Inevitably, therefore, it is duties against risk imposition - rather than fact-relative duties against harm - that are the proximate concern of a competent agent's moral decision-making. And if duties against risk imposition possess this central role in an agent's moral deliberation, it makes a certain degree of sense that they should possess a similarly central role in fixing the agent's moral liabilities - or, at least, those forms of liability that function as surrogates for compliance with the agent's moral duties.

There is, however, another question in the vicinity which is no less troubling. All of the cases discussed above - and all of the remaining cases I will discuss in this article - involve agents who are acting in a cognitively and epistemically flawless manner. These agent's beliefs are entirely consistent, therefore, with their evidence. But an adequate account of this moral domain will ultimately need to deal with threatening agents whose beliefs and evidence diverge. Consider, for example,

> *Altruistic Injection**: As in Altruistic Injection. However, Threatener has excellent evidence that Victim is genetically immune to the virus, such that injecting him with the antidote will kill him.

[50] See A. Slavny and T. Parr, "What's Wrong With Risk?", *Thought* 8 (2019) 2: 76 - 85. For a view on which there is somewhat greater distance between the morality of harming and the morality of risking harm, see J. Oberdiek, *Imposing Risk* (Oxford: Oxford University Press, 2017).

> Because Threatener is cognitively irrational, however, she believes that Victim is very likely *not* immune to the virus, and that injecting him with the antidote will thus very likely save him rather than kill him. Threatener is about to kill Victim by injecting him with the antidote. Observer can save Victim only by defensively shooting and killing Threatener.

If an agent's strict defensive liability were closely connected to her belief-relative obligations to internalize risk (and grounded in the stringency of her belief-relative duties against risk imposition), then Threatener would not be liable to defensive killing in this variant of Altruistic Injection: this variant would be morally on a par with the original case. But it seems to me that Threatener *is* liable to be killed here. This sort of judgment suggests that strict defensive liability is connected to evidence-relative duties rather than belief-relative ones. But I have no good explanation of why that should be so.

For these reasons, I have no fully worked out explanatory account of strict defensive liability to offer. Nevertheless, I believe, the picture of strict defensive liability that I have begun to develop in this section is a promising one. The principal aim of the following sections is not to further defend this picture. It is instead to intervene in a range of important debates about strict defensive liability. Each debate can be illuminated by thinking about the extensional connections - between duties against harm, duties against risking harm, obligations to internalize risk, and strict defensive liability - that I have highlighted in this section. I will argue that reflecting on these connections undermines the most prominent existing accounts of strict defensive liability. And in criticizing these existing accounts, I will often reference (by way of favorable contrast) the sort of explanatory account that I have begun to develop in this section. But even if this particular account is unsound, an adequate theory of defensive liability should ultimately prove consistent with and explain the relevant extensional connections. Alternatively, it must establish that the appeal of these extensional connections is illusory, and show that the intuitive judgments which these connections systematize should be rejected on reflection. Either way, relationships between strict defensive liability and adjacent moral phenomena such as obligations to internalize risk warrant much more philosophical attention. The critical work of the next several sections is, in part, an attempt to exhibit the analytical fertility of thinking about these relationships.

II. MORAL RESPONSIBILITY AND DISTRIBUTIVE FAIRNESS

The most prominent theory of strict defensive liability in the current philosophical literature is the *moral responsibility account*.[51] According to this view, an agent is liable to defensive harm when and because he is morally responsible for a fact-relative impermissible threat of harm to another person. This is true just in case it was reasonably foreseeable to the agent that his action might result in such harm.[52] For then the agent is more responsible than his victim for the fact that one of them must unavoidably suffer harm; thus it is fair that *he* should be the one to suffer harm, as opposed to his victim. And so, the moral responsibility account maintains, such a threatener forfeits his ordinary moral rights against the relevant harm's infliction.

The moral responsibility account has won many adherents, but it has also attracted considerable criticism. Much of this criticism has focused on the account's implications for cases such as Conscientious Driver. It is implausible, critics maintain, to suppose that conscientious driving (of the normal kind, that imposes only small and normal risks of harm) can render an agent liable to severe defensive harm. I believe this criticism is misguided. I have already argued that unusually risky acts of conscientious driving, as in the Sickly Driver cases, can subject a threatening agent to strict defensive liability. Soon I will argue that, at least in certain circumstances, normally risky conscientious driving can do so as well.[53]

The moral responsibility account should nevertheless be rejected, in my view, on other grounds. For its core structural features generate seriously counterintuitive implications, of a kind that has largely escaped critical attention. Consider, to begin,

> *Unowned Vial:* Threatener has strong reason to believe that he must consume a certain unowned vial of medicine, which is lying on the ground of an abandoned former hospital, if he is to survive a

---

[51] An account of this general kind was first suggested in K. Draper, "Fairness and Self-Defense," and M. Otsuka, "Killing the Innocent in Self-Defense." It has been developed most fully by Jeff McMahan. See, for example, *Killing in War* (Oxford: Oxford University Press, 2009); "The Moral Basis of Liability to Defensive Harm"; "Self-Defense Against Justified Threateners." Recently, Kerah Gordon-Solomon has offered an interesting variant of the account, in "What Makes a Person Liable to Defensive Harm?", *Philosophy & Phenomenological Research* 97 (2018) 3: 543 - 67. Gordon-Solomon dispenses with the condition that a faultless threatener's action be fact-relative wrongful, and shores up the view's luck egalitarian foundations. I believe that my criticisms apply to this version as well, but do not focus on it. Many other philosophers, including some of those cited on p. 2 fn. 5, above, hold similar views.

[52] The moral responsibility account is also intended to apply to culpable threateners, but my formulation and discussion of this view can safely leave this aspect of it aside.

[53] See section four below, at 52 – 73.

> deadly virus that he has just contracted. As he knows, there is some small chance that he is genetically immune to the disease's effects, such that using the medicine will be pointless. Threatener also knows that, if he uses the vial, it will be unavailable for use by a certain stranger, Victim, who has contracted the same deadly disease and is thus equally likely to die unless *he* uses the same vial. In fact, it happens that Threatener *is* genetically immune to the disease's effects, while Victim is not. In fact, therefore, only Victim has any use for the vial. Impartial Observer knows this, but she cannot communicate with Threatener before he uses the medicine. So Observer can prevent Threatener from using up the medicine only by shooting and killing him with her sniper rifle before he does.

It is clearly impermissible, it seems to me, for Observer to kill Threatener. Permissibly using up a scarce medical resource, on the basis of compelling but regrettably inaccurate medical evidence, does not make a person liable to be defensively killed. But the moral responsibility account implies otherwise. By using up the medicine, Threatener will be causally responsible for Victim's death (via double prevention).[54] And since there is in fact no reason for Threatener to use the medicine, in light of his genetic immunity to the virus, this action is fact-relative impermissible. So Threatener is responsibly and foreseeably generating a risk of fact-relative impermissible lethal harm to Victim, which is now about to eventuate. Thus the moral responsibility account implies that Threatener is liable to defensive killing.

The defender of the moral responsibility account might resist the criticism, as follows. Although Threatener is threatening to cause fact-relative impermissible lethal harm to Victim (via double prevention), he is not threatening to *do* harm to Victim. On some analyses of the moral asymmetry between doing and allowing harm, depriving another person of an unowned life-saving resource, and thereby enabling some pre-existing lethal process to harm him, does not count as doing harm to him.[55] And the moral responsibility account's defenders might now reformulate their view, so that it deems a faultless threatener liable to

[54] On causation via double prevention, see L.A. Paul and Ned Hall, *Causation: A User's Guide* (Oxford: Oxford University Press, 2013): 215 - 232.
[55] See, for example, J. McMahan, "Killing, Letting Die, and Withdrawing Aid," *Ethics* 103 (1993) 2: 250 - 279.

defensive harm only when he threatens to do harm, not just to cause it. But construing the moral responsibility account in this way is dubiously compatible with its underlying rationale. Even if Threatener is not threatening to do harm to Victim in Unowned Vial, she is more responsible than Victim for the threat of death which he faces, and for the fact that one of them must suffer lethal harm. Unlike Victim, Threatener has responsibly decided to run a risk that this very kind of unfortunate situation will eventuate: it is Threatener's "own voluntary action that has made it inevitable that one of them should die."[56] According to the moral responsibility account, that is the essential predicate of strict defensive liability.

Moreover, even if the proposed reformulation of the moral responsibility account secures its extensional adequacy in Unowned Vial, it fails to address the deeper issue the example brings to light. For the same basic problem arises even in cases in which a faultless threatener threatens to do harm to his victim. Consider the following, admittedly exotic case:[57]

> *Threatening Organ Donor:* Victim desperately requires an organ transplant, without which she will die in a few days. The only available organ belongs to a stranger, Threatener. It is permissible for Threatener to refrain from yielding up his organ, in light of the significant life-long burdens that doing so will impose on him. But Threatener has nevertheless decided, heroically, to provide the organ to Victim. Both parties are currently lying in the surgical theater, with their bodies cut open; soon a surgeon will transplant Threatener's organ into Victim. But now Threatener starts to suffer from a rare, deadly asthma attack; in all likelihood he will soon die, unless he immediately uses his albuterol inhaler. Unfortunately, as Threatener and Victim have just learned, Victim is seriously allergic to albuterol. For this reason, since using the inhaler will unavoidably release some albuterol particles into the surrounding air, there is a small chance that Threatener will kill Victim by using the inhaler. Moreover, Threatener knows there is some small chance that releasing the gas is fact-relative impermissible: there is some chance his respiratory ailment is just a temporary problem,

[56] McMahan, *The Ethics of Killing,* 402.
[57] The case is inspired by Judith Jarvis Thomson's famous violinist case and important variations on that case discussed by Frances Kamm. See J.J. Thomson, "A Defense of Abortion," *Philosophy and Public Affairs* (1973); F.M. Kamm, *Creation and Abortion* (Oxford: Oxford University Press, 1992).

such that releasing the gas is in fact entirely unnecessary. If that is true, and the gas will also in fact kill Victim, then it is fact-relative impermissible for Threatener to use the inhaler. An impartial Observer sees the situation unfolding; she knows that the Threatener's respiratory ailment is indeed a temporary problem, and that by using his inhaler he will in fact kill Victim. So Observer can save Victim only by killing Threatener (in which case Threatener's organ will still be available for transplantation into Victim.)

It seems implausible, to me at least, that Observer may defensively kill Threatener in order to save Victim. But the moral responsibility account implies otherwise, even on the reformulation under consideration. For Threatener is undeniably about to kill Victim. And Threatener is morally responsible (in the relevant sense) for this threat of impermissible harm. After all, he is quite aware of the possibility that his action might kill Victim - and do so impermissibly (because needlessly) relative to the facts. So the moral responsibility account deems Threatener liable to defensive killing.[58] But this is, it seems to me, an implausible result.

Defenders of the moral responsibility account might simply accept these results, of course. But reflecting on these examples in light of the presumptive parity principle and the risk internalization principle casts doubt on any such response. Let us begin with Unowned Vial. The fact-relative duty against bringing about another person's death, by using some unowned resource which he needs in order to save his life, is not nearly as stringent as the duty against killing another person by intruding on his body (for example, by hitting and killing him with a car, or shooting and killing him with a gun). An agent need not refrain from infringing the former sort of duty at the cost of her own life. It seems fact-relative permissible, for example, for one shipwrecked sailor to outswim another in order to take possession of, and thereby keep afloat by

[58] Defenders of the moral responsibility account might protest that Victim's pre-existing illness, rather than Threatener's choice to use his inhaler, has "made it inevitable that one of them should die." But *both* factors are causally responsible for this state of affairs. Defenders of the moral responsibility account might next point out that, although Threatener's choice to use his inhaler has made it inevitable that one of them should die, this is only because Threatener has previously agreed to provide his organ to Victim. So Threatener's entire course of action, viewed as a whole, has not made it inevitable that one of them should die. True enough. But even if this were not true, Threatener would not be liable to defensive killing. (Imagine that Threatener is in the surgical theater, and his organ is about to be transplanted into Victim, because the Society of Compassionate Consequentialists has kidnapped him and placed him there.)

holding onto, the only plank of wood in the vicinity.[59] Similarly, it seems fact-relative permissible for an agent to take and use an unowned vial of medicine, in order to save his life - even if he will thereby pre-empt another person's attempt to use the medicine herself, so bringing about her death (via double prevention).

If such fact-relative duties have only limited stringency, as I believe, then the corresponding evidence-relative duties against risk imposition will have only limited stringency as well. Thus, in Unowned Vial, Threatener will be not obligated to accept any given risk of death, in order to refrain from using the vial and thereby imposing an equivalent risk of death on Victim (via double prevention). This is true despite the fact that, should such a risk eventuate, Threatener's action would turn out to be fact-relative impermissible. And if Threatener need not internalize the risk of death that he is imposing on Victim, in this way, then he is not liable to defensive measures that will in effect force him to internalize this risk. Taken together, therefore, the presumptive parity principle and the risk internalization principle correctly imply that Threatener is not liable to be killed in Unowned Vial.

The same analysis applies to Threatening Organ Donor. As I have said, the fact-relative duty against killing another person by intruding on his body is presumptively a very stringent one. And Threatener is threatening to infringe such a duty in this case. But something about the situation - such as the fact that, by killing Victim, Threatener would be depriving him of a good (continued life) that Victim can enjoy only through the use of Threatener's body - seems to significantly reduce the ordinary stringency of this fact-relative duty.[60] In consequence, it seems, Threatener's fact-relative duty against killing Victim has only limited stringency in these circumstances. Thus Threatener would not be fact-relative obligated to let himself die, or impose death on himself, rather than infringe this duty. If that is correct, then Threatener's evidence-relative duty against imposing any given risk of death on Victim will *also* have only limited stringency in these circumstances: Threatener will not be evidence-relative obligated to accept such a risk of death, or to impose it on himself, rather than impose it on Victim. Under the risk internalization principle,

[59] See also the discussion of such cases below at pp. 61 – 63.
[60] See Kamm, *Creation and Abortion,* esp. pp. 22 - 33.

therefore, Threatener is not liable to a defensive measure that will, in effect, force him to internalize such a risk. So he is not liable to be defensively killed, in order to prevent the risk of death that he is imposing on Victim from materializing.

The moral responsibility account struggles with these cases because its attributions of defensive liability are insensitive to the stringency of the underlying duties that a faultless threatener is about to infringe. Instead, the moral responsibility account maintains, the case for strict defensive liability sounds in *distributive fairness*: responsibly posing a threat of fact-relative impermissible harm renders a faultless threatener liable to defensive harm because, if either he or his victim must suffer harm, it is more fair that he should be the one to do so.[61] For the threatener's own responsible choices have brought about the tragic situation at hand. But in both Unowned Vial and Hook-Up, the only party who has made any responsible choice, thereby responsibly bringing about the situation at hand, is the threatener. The victim has made no such choice at all (or so we may easily stipulate).[62] The logic of the moral responsibility account thus implies that it is fairer for the threatener to die than his victim, such that the threatener is liable to be defensively killed. But that is implausible, or so I have claimed.

There is a tempting line of defense to which proponents of the moral responsibility account might now retreat. The conception of fairness on which the moral responsibility account relies is a broadly luck egalitarian one. The guiding idea is that, if someone makes a suitable gamble and his gamble goes awry, it is fair for him to suffer the resulting consequences.[63] But as many luck egalitarians have recognized, not all

---

[61] For related and complementary criticisms of the moral responsibility account's appeal to distributive fairness, see Quong, *The Morality of Defensive Force,* e.g. at p. 91; Burri, "Morally Permissible Risk Imposition." Burri acutely observes that choices to impose risk will only bear on (luck egalitarian) distributive fairness, in the way the moral responsibility account requires, against the backdrop of a suitable initial distribution of resources - a distribution of a kind that, in the real world, often fails to obtain. She is tempted to draw the lesson that strict defensive liability arises much less often than we might initially suppose. In my view, the lesson is instead that strict defensive liability should not be connected to anything like luck egalitarian distributive fairness.

[62] We may stipulate that Victim's need for the vial in Unowned Vial is entirely the result of a genetic susceptibility to the virus, for which he bears no responsibility. And we may similarly stipulate that, in Threatening Organ Donor, Victim has been hospitalized (and so put in the position that leaves him susceptible to being killed by Threatener's evidence-relative permissible but fact-relative impermissible choice to use the albuterol inhaler) because of a medical malady that was foreordained by his genetic constitution.

[63] See G.A. Cohen, "On the Currency of Egalitarian Justice," *Ethics* 99 (1989) 4: 906 - 944; K. Lippert-Rasmussen, *Luck Egalitarianism* (London: Bloomsbury, 2015). The connection between luck egalitarian distributive fairness and the

gambles can plausibly play such a role. If I walk to the pharmacy to purchase my monthly medications, and I am paralyzed by a freak bolt of lightning along the way, the resulting state of affairs is not a fair one. Some line must be drawn between this sort of choice and, for example, the choice to gamble one's life savings in a poker game (the luck egalitarian paradigm of a responsible decision whose consequences fairly lie with the agent who makes it). And supposing that luck egalitarians can successfully draw the needed line, it seems clear that choices to take evidently necessary and entirely reasonable measures to save one's life will fall decidedly on one side of it. Such choices are not relevantly like gambling one's life savings in poker.

Imagine, for example, that a villain credibly threatens to murder an innocent person unless he plays Russian roulette on himself (*Self-Preserving Russian Roulette*). The innocent gets unlucky and inflicts lasting brain damage on himself. It is not at all fair for this person to suffer the consequence of his responsible choice, any more than if this harm had been visited on him by a freak bolt of lightning.[64] Any plausible interpretation of the basic luck egalitarian idea will not imply that such harmful consequences are fair. For the same reason, it can be argued, the basic luck egalitarian idea (plausibly interpreted) will not imply that it is fair for the faultless threateners in Unowned Vial and Threatening Organ Donor to bear the harmful consequences of *their* actions. For these characters are also taking evidently necessary and reasonable measures to preserve their lives. And so the moral responsibility will not - when its luck egalitarian foundations are properly understood - saddle these threateners with defensive liability.

The problem is that this argument generalizes unacceptably. After all, no less than the threateners in Unowned Vial and Threatening Organ Donor, the threatener in Mistaken Attacker is taking an evidently necessary and fully reasonable measure to preserve his life. But that case is among the most compelling putative instances of strict defensive liability; indeed, it is largely by reference to such cases that the moral responsibility account was initially articulated and defended. Abandoning the intuitive judgment that the

---

moral responsibility account is most fully embraced and developed in Gordon-Solomon, "What Makes a Person Morally Liable to Defensive Harm?". The connection is also explicitly made in Otsuka, "The Moral Responsibility Account."

[64] It might be suggested that the agent's choice is not in fact voluntary or responsible, because he is being coerced. I doubt that coercion undercuts responsibility in this way. If it does, however, the case may be reconstructed such that the agent is acting in response to a risk of being killed by a natural threat.

threatener in this case is liable would thus constitute a serious blow to the moral responsibility account's explanatory aspirations. To develop the point, it will help to have a more fully described version of the case in our view:

> *Mistaken Attacker:* Threatener is walking around the town square when he receives an alert on his cell phone, from the police. The alert notifies him that a notorious knife-wielding murderer has escaped from prison, and is attempting to murder one randomly chosen innocent person in the next town he visits. The alert contains a photo of this notorious murderer, which exhibits the murderer's distinctive facial structure. When Threatener looks up from his phone, he is startled to realize that there is a man proceeding towards him. This man, Victim, looks just like the escaped murderer and appears to be brandishing a knife. Threatener reasonably concludes that Victim almost certainly *is* the escaped murderer, and that he can save himself from being murdered only by using his protective firearm to shoot and kill Victim in self-defense. As Threatener knows, there is some chance that Victim is a wholly innocent and unthreatening person. But he believes, fittingly on his evidence, that this chance is very small. So he pulls out his firearm and prepares to use it. In fact it turns out that Victim *is* a wholly innocent and unthreatening person. He is the knife-wielding murderer's identical twin, who is completely unaware of his murderous brother's existence; the knife he appears to be holding is, in truth, his wallet. An impartial third party, Observer, happens to be nearby, and she knows all of the relevant facts about the situation. She sees Threatener preparing to attack Victim, and realizes that she can save Victim only by shooting and killing Threatener before he shoots and kills Victim.[65]

If we say (quite plausibly) that an agent's choice to take such an evidently necessary measure to preserve does not render it fair for him to suffer the resulting harm - and we follow the moral responsibility account in connecting defensive liability to distributive fairness - then we are committed to denying that the threatener in this case is liable to be killed. But it is intuitively plausible that he *is* liable to be killed.

[65] This version of the case blends elements found in Otsuka's version, in "Killing the Innocent in Self-Defense," p. 91, and McMahan's version, in "The Moral Basis of Liability," p. 387.

No intuitive judgment is beyond revision, of course. We might conclude with chagrin that our commitments about distributive fairness compel us to abandon this judgment. But the risk internalization principle and the presumptive parity principle suggest that discarding this judgment in the name of fairness would be a mistake. To see why, consider

> *Mistaken Attacker With Risk-Internalizing Gun:* As in Mistaken Attacker. However, Threatener's firearm is a technologically sophisticated one, with an advanced facial recognition function that he can (if he wishes) enable before firing it. If Threatener enables this facial recognition function, uploads the picture of the serial killer (which he has received from police) from his phone into the gun, and subsequently pulls the gun's trigger, the gun will scan Victim's face and kill him if and only if it accurately determines that Victim is in fact the serial killer. If the facial recognition software instead determines that Victim is simply a look-alike of the serial killer, it will protectively jam the gun's chamber in response, preventing it from firing. However, if the gun does this, it will also blow up (because of a technological malfunction) and kill Threatener. Threatener knows all this. Thus he knows that, by enabling the gun's facial recognition function before he shoots at Victim, he can internalize the risk that he will be killing an innocent person by doing so. If Threatener does *not* want to put himself at risk of death, he can simply refrain from enabling the gun's facial recognition function before shooting it at Victim.

It seems obvious to me that Threatener must internalize this risk, by enabling his gun's facial-recognition function before he shoots it. This judgment can be supported, moreover, by appeal to the presumptive parity principle.[66] And if Threatener lacks any such mechanical means of risk internalization, but he *can* request a functionally equivalent course of defensive action from an appropriately situated impartial

[66] As follows. Fact-relative duties against killing non-liable, innocent people by intruding on their bodies - the sort of duty that Threatener is about to infringe - are very stringent. Thus an agent would be fact-relative obligated to save himself from a fate worse than death by shooting and killing himself, rather than by shooting and killing another innocent person. Evidence-relative duties against imposing risks of harm on non-liable, innocent people, by intruding on their bodies, are similarly stringent. Thus an innocent agent would presumptively be evidence-relative obligated to save himself from a large risk of death by playing Russian roulette on himself, rather than by playing Russian roulette on another innocent person. If this is correct, then Threatener is similarly obligated to use his gun's facial recognition function, thereby internalizing the risk that by shooting Victim he will be killing an innocent and non-liable person.

observer, he is obligated to do so. Similarly, the risk internalization principle implies, Threatener is liable to such defensive action at the hands of an impartial observer even if he is not actually in a position to request it. That is the situation in Mistaken Attacker.

The preceding argument helps to illuminate why Threatener is liable to defensive killing in Mistaken Attacker, in order to prevent him from killing Victim, although it is no fairer (in the luck egalitarian sense) that he should be the one to die.[67] Threatener's liability to defensive killing is no more grounded in luck egalitarian distributive fairness than the fact-relative duty against killing Victim from which his strict liability ultimately derives. This more fundamental moral phenomenon cannot be annulled or weakened in the name of fairness. An agent may not flip a coin and, if he wins the flip, proceed to shoot and kill another innocent person as a means or side-effect of saving his own life.[68] Similarly, it is in no way fair that an innocent agent should find himself in a position where he can save his life only by driving to the hospital, thereby hitting and killing another person with his car. The situation is no fairer than if the parties' positions were reversed. Yet the agent is still obligated not to hit and kill the latter. Nor is the stringency of this duty susceptible to weakening in the name of fairness: the agent cannot flip a coin and proceed to the hospital if it lands in his favor.

If our fact-relative duties against harm are indeed insensitive to luck egalitarian distributive fairness, in this way, we should expect that this insensitivity will be transmitted to the derivative moral phenomena - such as evidence-relative duties against imposing risk, obligations to internalize risk, and strict defensive liability - which these fact-relative duties ground. And that is indeed what we find. It is no more fair, in Mistaken Attacker With Risk-Internalizing Gun, that Threatener should be subject to a small risk of death than Victim. Nonetheless, Threatener is obligated to internalize the risk of death he is creating, rather than

[67] If anything, fairness would seem to require that Observer to give Threatener and Victim an equal chance at living, by (say) flipping a coin and defensively killing Threatener only if the coin flips heads. But this seems incorrect - just as it seems incorrect that, in the risk-internalizing gun variant of the case, Threatener should be permitted to flip a coin and use the gun's facial recognition function (thereby internalizing the risk he is imposing on Victim) only if he loses the flip.

[68] Similarly, it is impermissible for an agent to flip a coin and, depending on how it lands, take another person's vital organs in order to save his own life. Quong, *The Morality of Defensive Force,* p. 91; Otsuka, "Killing the Innocent in Self-Defense," p. 76 fn. 9. See also Walen, *The Mechanics of Claims,* pp. 60 - 62.

impose it on Victim. And Threatener is liable to defensive measures that will in effect force him to internalize this risk, even if (as in the original Mistaken Attacker) he cannot choose to internalize it.

The same analysis applies, I believe, to Sickly Driver and many other cases: in each case it is no fairer that a faultless threatener should die than his victim, and yet the former is liable to be defensively killed. Perhaps there are other cases in which luck egalitarian considerations *do* have some bearing on defensive liability. If someone keeps a pet tiger for pleasure, taking all due care to prevent it from mauling anyone, perhaps it is fair that he should be mauled by the tiger (if anyone is).[69] The choice to keep a tiger may well be analogous to the choice to gamble one's life savings at a casino. If so, the luck egalitarian conception of fairness may *strengthen* the case for strict defensive liability in such examples. But the core case for strict defensive liability sounds in the stringency of our evidence-relative duties against imposing risk (which in turn derive from our fact-relative duties against harm), not in luck egalitarian fairness.

Indeed, there is an important sense in which this vision of our moral duties and liabilities is an *unfair* one: the relevant duties and liabilities, when respected and enforced, can render some people much worse off than others through mere luck, rather than any fault or choice of their own. Philosophers who reject strict defensive liability have been moved, in part, by the fact that strict defensive liability is evidently unfair in this way. But this is, as we have seen, a kind of unfairness that is pervasive in deontological morality. And while this tension between distributive fairness and deontological morality is a troubling one, its very pervasiveness and generality suggests that we should not abandon the relevant convictions about strict defensive liability unless we are prepared to abandon the wider edifice of deontological moral thought. For our convictions about strict defensive liability inherit their tension with luck egalitarian distributive fairness from the more basic deontological moral phenomena - concerning the stringency of fact-relative duties against imposing harm, and evidence-relative duties against imposing risks of harm - from which strict defensive liability derives.[70]

---

[69] This case is a bare-bones version of Tiger Owner, discussed below at pp. 56 – 58.

[70] Other dimensions of the general tension between deontological morality and distributive fairness are observed in G.A. Cohen, *Self-Ownership, Freedom and Equality* (Oxford: Oxford University Press, 1995): 243 – 244; and S. Scheffler, *Boundaries and Allegiances: Problems of Justice in Liberal Theory* (Oxford: Oxford Univeristy Press, 2001): 73 – 75.

This leaves open the possibility, however, that strict defensive liability is partly grounded in (or constitutively connected to) some *distinctively deontological* conception of fairness. Consider

> *Reciprocal Russian Roulette:* In order to save himself from a near certainty of death, Tomas must impose a 1% risk of death on Victoria, by shooting at her with a revolver that contains a hundred chambers (one of which is loaded). He has no way to internalize this risk. Meanwhile, in order to save *her* life from a near certainty of death, Victoria must impose a 1% risk of death on Tomas, by shooting just such a revolver at him. Victoria *can* internalize this risk, by enabling a feature on her gun that will blow it up (killing her) if and only if pulling the trigger is about to discharge the one loaded chamber (thereby killing Tomas).

It seems to me, intuitively, that Victoria is not obligated to internalize the risk she is imposing on Tomas, and that she is not liable to defensive measures that will in effect force her to internalize this risk.[71] This judgment might be explained and defended in various ways.[72] But one natural thought is that it is *unfair* that Tomas should take a gamble with Victoria's rights that is greater than the gamble she is taking with his rights. The unfairness of such a relationship might explain why Victoria may permissibly take an equivalent gamble with Tomas's rights, by imposing an equivalent risk of death on him, and why she is not liable to defensive measures that will in effect force her to internalize this risk.

In this way, obligations to internalize risk and defensive liability may be grounded in or constitutively connected to a certain, distinctively deontological conception of relational fairness: a conception that is essentially concerned with interacting agents' equal or reciprocal compliance with their directed moral duties to one another. Even if this is correct, however, the existence of these very moral duties will often offend against other, more familiar conceptions of fairness (such as the luck egalitarian

---

[71] To be clear, I am not here claiming that this sort of judgment is defensible on reflection. See the discussion of the reciprocal risk account below at pp. 63 - 66, and the further discussion in "Strict Liability to Defensive Harm II," pp. 109 – 125.

[72] *Id.*

one).[73] Or so I have argued. If I am correct, the moral responsibility account goes awry in explaining strict defensive liability by reference to any such conception. But it is equally an error to reject strict defensive liability for lack of fairness, unless we are prepared to reject basic deontological moral commitments about the stringency of our duties against harm.

III. The Moral Status Account and Agent-Relative Prerogatives to do Harm

Perhaps the most prominent alternative to the moral responsibility account in the current philosophical literature is Jonathan Quong's *moral status account* of liability to defensive harm. According to this view,

> A is liable to defensive harm for ϕ-ing when: (a) the evidence-relative permissibility of ϕ-ing depends on the assumption that at least one person, B, lacks a moral right, but (b) B in fact possesses the relevant moral right, and thus (c) B faces a threat, or apparent threat, to her rights.[74]

Thus, for example, Threatener is liable to be killed in defense of Victim in Mistaken Attacker. It is because Threatener has evidence that Victim is a serial killer, attempting to murder her, that her attack on him is permissible. Her attack would be impermissible if she lacked this evidence. In this sense, the evidence-relative permissibility of Threatener's attack on Victim depends on the assumption that he is liable to defensive killing.[75] Thus it is only fair that Threatener should lose her own rights against being killed (as a defensive measure, in order to save Victim's life) when this assumption proves to be mistaken. Or so the moral status account maintains.

---

[73] In fact, I believe, there is a similar tension between deontological morality (strict defensive liability included) and distributive fairness even if we understand fairness exclusively as a social, political, or institutional moral ideal. After all, political institutions and social practices may subject a person to an unjustly high risk of finding himself in a situation where he is constrained by stringent moral duties against harming other people. For example, if there is injustice in the background allocation of his society's medical resources, then Threatener in Sickly Driver might have been the victim of an unjustly high risk of finding himself in precisely such a tragic situation. But this does not, I think, diminish the stringency of his fact-relative duty against hitting and killing Victim with his car. Nor does this background unfairness seem to diminish the stringency of Threatener's evidence-relative duties against imposing risks of such harm on Victim. It is not hard to imagine that, if Threatener is allowed to drive to the hospital without internalizing the risk of death he will thereby impose on Victim, or incurring strict defensive liability for doing so, the justice of the overall distribution of risks and harms in society will be enhanced. This will be so if Victim has been provided, by her society and state, with an unjustly high share of security from bodily harm (as compared to level of security provided to other people, such as Threatener). Yet Threatener will still be obligated to internalize the risk of harm he is imposing on Victim, and he will still be liable to defensive measures that in effect force him to internalize this risk.

[74] Quong, *The Morality of Defensive Force,* p. 38.

[75] Quong clarifies that this is the relevant sense of dependence in *The Morality of Defense Force,* pp. 94 – 95.

By contrast, the moral status account implies, Threatener is *not* liable to defensive killing in Conscientious Driver. For the evidence-relative permissibility of his threatening action (driving himself to the cinema, in the vicinity of pedestrians) does not depend on the assumption that any pedestrian is liable to harm. It depends on the assumption that driving to the cinema will impose only a small risk of harm on nearby pedestrians. The moral status account also implies, on similar grounds, that the threatener in Sickly Driver 5 is not liable to defensive harm. For the evidence-relative permissibility of Threatener's decision to drive himself to the hospital, thereby imposing a 1% risk of death on Victim, does not depend on the assumption that Victim lacks his ordinary rights against being harmed.

I have argued, of course, that the threatener in Sickly Driver 5 *is* liable to lethal defensive harm. Insofar as this position is plausible, and my risk internalization argument for it is persuasive, the moral status account is open to objection. There is a sense, however, in which this objection is dialectically inapposite. For my risk internalization argument depends on the assumption that the duty against hitting and killing another person with a car is a highly stringent one - sufficiently stringent that an agent is presumptively obligated to let himself die rather than infringe this duty. And there is a significant minority of philosophers writing on the morality of defensive harm, Quong among them, who would contest this assumption. These philosophers recognize an *agent-relative prerogative to do harm:* they believe that an agent is permitted to inflict (a risk of) harm on another innocent person in order to spare himself from (a comparable risk of) comparable harm, provided that the agent does not thereby use the other person's body or property as a means.[76]

If this philosophical view is correct, then the fact-relative duty against killing another innocent person is not so stringent that an agent is fact-relative obligated to let himself die rather than infringe this duty. In Sickly Driver 1, for example, Threatener may permissibly hit and kill Victim with his car, in order to avoid suffering death himself. This view similarly implies that, if an agent can escape a very bad fate

[76] N. Davis, "Abortion and Self-Defense," *Philosophy & Public Affairs* 12 (1984) 3: 175 - 207; Quong, *The Morality of Defensive Force,* pp. 58 - 96. Larry Alexander and Kim Ferzan may also embrace such a view, in *Crime and Culpability: A Theory of Criminal Law* (Cambridge: Cambridge University Press, 2009): 136 - 141. But there is another plausible interpretation of their view, as Ferzan elsewhere observes. See K.K. Ferzan, "Self-Defense, Permissions, and the Means Principle," *Ohio State Journal of Criminal Law* 8 (2011) 503: 503 - 513, 504.

either by imposing a small risk of death on another innocent person or by imposing an equivalent risk of death on himself, he may (evidence-relative) permissibly do the former. If this is correct, then Threatener is *not* obligated to internalize the 1% risk of harm he is imposing on Victim by driving himself to the hospital in Sickly Driver 3 and Sickly Driver 4. Nor will Threatener be liable, under the risk internalization principle I have defended, to functionally equivalent defensive harm in Sickly Driver 5.

To be clear, philosophers who defend the agent-relative prerogative to do harm have not explicitly discussed risk internalization. But these positions are entailed by the basic structure of the prerogative they defend. And these positions are in direct tension, of course with the intuitive judgments on which I have relied. It may therefore seem that risk internalization arguments for strict defensive liability have nothing to say to these philosophers. But that is not true. Reflection on the morality of internalizing risk can put serious pressure on these philosophers' views, by revealing tensions between their central commitments.

Let us begin with Quong's views, in particular. In unpublished work,[77] I have argued that the moral status account struggles with cases of the following kind:

> *Mistaken Attacker**: Threatener has strong reason to believe that Victim is an (unknowing) sleepwalker, and that she is about to unconsciously strangle him. Threatener thus moves to kill Victim in self-defense. In fact, however, Victim is an entirely unthreatening person, who poses no threat to Threatener or anyone else.

Intuitively, it seems to me, Threatener is liable to be defensively killed - no less in the original Mistaken Attacker. But the moral status account implies otherwise. For the evidence-relative permissibility of Threatener's attack on Victim does not depend on the assumption that Victim is liable to (i.e. lacks a right against) any harm. It depends on Threatener's evidence that Victim is a non-liable threat to his life, such that he has an agent-relative prerogative to infringe her right against being killed.[78] The moral status

[77] K. Ramakrishnan, "Quong on Defensive Harm" (unpublished manuscript). The case I present in that manuscript ("*Mistaken Attacker II*") is very similar to, but not exactly the same as, the case I offer in the text above ("*Mistaken Attacker**"). Quong responds to the manuscript in *The Morality of Defensive Harm*, p. 92 - 95.

[78] After all, by the lights of the moral status account itself, Victim would possess a right against being killed even if she were about to strangle Threatener in her sleep.

account thus implies that Threatener is *not* liable to defensive harm, and that an impartial observer may not defensively kill him. In my view, this is a deeply counterintuitive result. It seems intuitively plausible to me that, by attacking Victim, Threatener loses his right against being killed in both cases: trying to shoot and kill another innocent person is, after all, a morally grave affair. But there is also an accessible frame of mind in which it strikes me as somewhat strange - in both cases - that Threatener's fully conscientious and reasonable attempt to save his life should deprive him of his right against being killed. There is no accessible frame of mind in which I am tempted to impute defensive liability in one case but not the other.

The dialectical force of this objection might be limited, however, for Quong has reported that he simply disagrees.[79] But I previously argued that Mistaken Attacker* reveals an even deeper problem for the moral status account: namely, it cannot explain why Threatener is liable to defensive killing *even in the original Mistaken Attacker*. In that case, of course, Threatener has evidence that Victim is a liable threat (rather than a non-responsible, non-liable threat, as in Mistaken Attacker*). But the permissibility of Threatener's attack on Victim does not depend *on the assumption that she is liable.* It depends only on the assumption that she is a threat to his life, liable or not. If Threatener were to learn that Victim is a non-responsible (and thus non-liable) threat to his life, his action would still be evidence-relative permissible, in light of his agent-relative prerogative to do harm. And so, I claimed, it is incorrect to suppose that the evidence-relative permissibility of Threatener's attack depends on the assumption that Victim is liable to harm, even in the original Mistaken Attacker case.

I am now persuaded by Quong that this further objection rests on a misinterpretation of his view.[80] The question, under the moral status account, is not whether Victim's evident liability (in Mistaken Attacker) plays an essential role in rendering Threatener's attack on her evidence-relative permissible. It is whether Threatener's attack on Victim would be permissible in the absence of *his particular factual evidence for thinking that she is liable*. Consider the closest possible world in which Threatener lacks evidence that Victim

[79] Quong, *The Morality of Defensive Force,* pp. 94 - 95.
[80] It is a misinterpretation that is arguably invited by some of Quong's official formulations of the moral status account, at e.g. *The Morality of Defensive Harm,* p. 38.

is a serial killer trying to murder him (such that she is liable to defensive harm). In this possible world, it seems plausible to say, he does not have evidence that she is threatening him at all.[81] In this sense, the evidence-relative permissibility of Threatener's attack on Victim depends on his evidence for thinking that she is a liable threat.

So clarified, the moral status account evades my objection that it cannot explain Threatener's defensive liability in Mistaken Attacker. But the view then succumbs to an even more damning problem: it makes a threatener's defensive liability depend, in bizarre ways, on irrelevant information. Consider

> *Blue Button:* Tara has excellent reason to believe that her neighbor Gerardo has poisoned her, and that she can save her life only by pressing a certain blue button. But there is a small chance, as she knows, that Gerardo is completely innocent and neither he nor anyone else is threatening her. If that is true, Tara knows, then pressing the blue button will randomly select and kill one innocent person in her town.

Suppose that Tara's evidence is misleading, and she is now (therefore) about to kill a random innocent person in her town by pressing the blue button. The moral status account implies that she is liable to be defensively killed, in order to save this innocent victim's life, if and only if he happens to be Gerardo. For the evidence-relative permissibility of Tara's choice to press the blue button (thereby risking harm to innocent people) depends on her evidence that Gerardo has culpably poisoned her, such that he is liable to harm. The permissibility of her action does *not* depend on any evidence concerning anyone else. But this conclusion - that Tara's defensive liability depends on whether the innocent person she is about to kill happens to be the particular person who figures in her misleading evidence about the apparent threat to her life - seems deeply implausible. An impartial observer need not ascertain whether Tara's innocent victim happens to be Gerardo in order to determine whether to deploy defensive force against her.

The point can be reinforced by thinking about risk internalization. Suppose that Tara can internalize the risk that she will be killing an innocent person by pressing the blue button. If Tara has a

[81] Of course, there are some possible worlds (such as the one envisaged in Mistaken Attacker*) in which Threatener *does* have evidence that he is under threat by Victim. But these possible worlds are farther away.

general agent-relative prerogative to inflict (a risk of) harm on another innocent person in order to spare herself from an equivalent (risk of) harm, then she is permitted to refrain from internalizing this risk. If there is no such prerogative - if duties against (risking) harm are highly stringent - then Tara *is* obligated to internalize this risk. The latter view is more plausible, in my view, but both positions make sense. Each view reflects a coherent conception of the stringency of our interpersonal duties and the scope of our agent-relative prerogatives to harm. By contrast, it makes little sense to suppose that Tara must internalize the risk that she will be killing the innocent Gerardo but need not internalize the risk that she will be killing any other innocent person. It seems irrelevant, in determining whether Tara must internalize these various risks of death, that one of the innocent people she might be killing happens to figure in the evidence that justifies her imposition of these risks. What seems relevant is the stringency of Tara's duties against killing innocent people; she is obligated to internalize all of these risks, or none of them. Similarly, I believe, she is liable to lethal defensive harm on behalf of each innocent victim to whom she is risking death, or else she is liable on behalf of none of them.

If this is correct, Quong's moral status account delivers implausible verdicts about defensive liability even on the assumption that there is an agent-relative prerogative to inflict (risks of) harm on other innocent people. But there are other philosophers who recognize such a prerogative as well. And some of these philosophers argue that *no* faultless threateners are liable to defensive harm. This blanket rejection of strict defensive liability is not impugned by the argument I have just pressed against Quong. And my previous risk internalization arguments for recognizing the existence of strict defensive liability may *also* seem powerless against this view. For my arguments have relied on intuitive judgments about the considerable stringency of duties against (risking) side-effect and eliminative harm. If there is an agent-relative prerogative to inflict such harm, however, then the relevant duties are not as stringent as I have supposed. And so, it may appear, the blanket rejection of strict defensive liability cannot be challenged by such arguments.

That is not so, however. For virtually all nonconsequentialist moral philosophers affirm the considerable stringency of moral duties against *manipulative harming:* harmfully using another person's body

without his consent.[82] And we can construct a risk internalization argument which exploits the stringency of these duties, in order to establish the existence of strict defensive liability in a narrower but philosophically significant set of cases.

To begin, suppose that an innocent agent, Threatener, can save herself from a malicious assailant's bullets by grabbing hold of an innocent bystander, Victim, and using his body as a human shield, thereby killing him (*Manipulative Attacker 1*).[83] This is impermissible, in both the fact-relative sense and the evidence-relative sense: Threatener's duty against manipulatively killing Victim is sufficiently stringent that she must allow herself to die rather than infringe this duty. There is no agent-relative prerogative, in other words, to save one's life by lethally using another person's body as a protective shield.

Suppose, however, that Threatener has excellent reason to believe that her use of Victim as a shield will not in fact harm him.[84] If (for some such reason) the risk of harm that Threatener will be imposing on Victim is sufficiently low, it is evidence-relative permissible for Threatener to use Victim's body in this way (*Manipulative Attacker 2*). Just how low the risk must be is far from clear, of course. But it is hard to deny that an agent may impose *some* sufficiently small risk of death on another person, by non-consensually using his body as a means, in order to save herself from a near certainty of death.

Next suppose that Threatener can internalize the small risk of death that she is permissibly imposing on Victim by non-consensually using his body: she can activate a risk-internalizing mechanical device that will monitor the situation and defensively kill her if and only if her use of Victim is about to kill him (*Manipulative Attacker 3*). It seems clear that Threatener is obligated to internalize this risk. (Compare: if an agent can save her life from a deadly disease by using her own body in a potentially harmful way - e.g. by running some potentially deadly medical tests on herself - or by using another person's body in a potentially harmful way - e.g. by running some potentially deadly medical tests on her identical twin - the agent is obligated to save her life in the former way.)

---

[82] On the concept of manipulative harming, see above p. 6 fn. 12.
[83] For a similar case, see Otsuka, "Killing the Innocent in Self-Defense," p. 76.
[84] Perhaps Victim happens to be the assailant's favorite movie star, whom he adores, and whom he would never try to harm.

Suppose next that Threatener can internalize this risk by communicating a request to an impartial observer, who can defensively monitor the situation and shoot Threatener (killing her) if and only if the observer reliably determines that Threatener's use of Victim's body is about to kill him (*Manipulative Attacker 4*). Here too it seems clear that Threatener is obligated to internalize the risk. She may not risk killing Victim, by using his body for her own benefit, when she can instead procure the same expected benefit by imposing an equivalent risk of death on herself. And if that is correct, it is also plausible to suppose that Threatener is liable to be defensively killed (if necessary) by the impartial observer, in order to prevent her from killing Victim by using his body, whether or not Threatener is aware of the observer's presence or able to communicate with him (*Manipulative Attacker 5*). After all, Threatener would be obligated to internalize this very risk, if she could, by asking the observer to engage in precisely this defensive course of action.

This risk internalization argument is obviously compressed, as compared to the risk internalization argument about the Sickly Driver cases developed in section one. But the structure of these arguments is exactly the same. Moreover, both arguments draw support from the plausible extensional principles (the presumptive parity principle and the risk internalization principle) that I developed in that section. So I believe that this new risk internalization argument establishes that at least some faultless threateners - namely, threateners about to manipulatively kill their victims – are indeed liable to lethal defensive harm. If I am correct, any wholesale rejection of strict defensive liability is difficult to defend. So long as a philosopher acknowledges that duties against manipulative harming are highly stringent, as virtually all nonconsequentialist moral philosophers do, she is under considerable reflective pressure to accept that agents who threaten to infringe such duties do indeed become liable to equivalent defensive harm.

Nevertheless, since the relevant philosophers deny that duties against *non-manipulative* harming are highly stringent, they can coherently deny that faultlessly threatening non-manipulative harm (as in Sickly Driver, Conscientious Driver, Tiger Owner, Mistaken Attacker, and many other cases) incurs strict defensive liability. My own view, of course, is that many duties against non-manipulative harming *are* highly stringent. It seems to me impermissible for an agent to save his life by driving to the hospital if he will unavoidably hit and kill another innocent person along the way; or to save himself from a deadly disease by

releasing and inhaling a medical gas which will then seep into his neighbor's house, killing her; or to save himself from a deadly projectile by detonating a grenade that will kill another person in the vicinity; or to save his life by shooting a gun in the direction of a crowd, killing someone in it. In all of these cases, it is impermissible for an agent to harm another person, even as a mere side-effect, by intruding on her body. To be sure, there may be other cases in which the duties against (risking) harmfully intruding on another person's body lack their ordinary stringency. That is plausibly true in Threatening Organ Donor.[85] And it is plausibly true in Altruistic Injection, as well as several other cases (which we will later discuss) in which a faultless threatener is acting in order to benefit other people, rather than himself. But across a significant range of cases - including most cases in which an agent is acting entirely for his own benefit, with an entirely self-regarding justification - an agent's duties against (risking) harmfully intruding on another person's body will be very stringent indeed. Or so I believe, in common with most other nonconsequentialist philosophers writing on the morality of defensive harm.

Nevertheless, I must concede, I have no decisive argument against the opposing view, according to which duties against non-manipulative harm are much less stringent than I suppose. And the opposing view is not entirely unreasonable. It is not entirely unreasonable, therefore, to deny that strict defensive liability exists in as wide a range of circumstances as I would maintain. In the end, I think, our conclusions about an agent's obligation to internalize the risk of harm that he is imposing on another person (and his liability to suffer defensive harm aimed at preventing this risk from eventuating) will be inextricably bound up with our commitments about the stringency of the primary duties against harm that he is threatening to infringe. That is unsurprising if - as the presumptive parity principle and the risk internalization principle maintain - obligations to internalize risk and strict defensive liability are closely connected to the stringency of evidence-relative duties against risk imposition, which are in turn closely connected to the stringency of fact-relative duties against harm. If these extensional connections are plausible, disagreement and uncertainty

[85] See above, pp. 33 – 34.

about the stringency of duties against harm will predictably generate (or find expression in) disagreement and uncertainty about strict defensive liability and obligations to internalize risk.

IV. The Varieties of Conscientious Driving

I have argued that we should reject both the moral responsibility account and the moral status account. But it is independently important to determine whether we should accept the moral responsibility account's most famous and controversial commitment. Is it implausible to suppose that a conscientious driver who threatens to kill a pedestrian, through an ordinary and minimally risky act of driving, is liable to be defensively killed? Is it implausible, in other words, to deem the threatener in Conscientious Driver liable to lethal defensive harm?

The question is important not only because driving is ubiquitous in modern industrial societies. It is also important because, to a significant extent, philosophical arguments against recognizing strict defensive liability in Conscientious Driver have doubled as philosophical arguments against recognizing strict defensive liability more generally. That is not universally true; Quong has argued that we should reject strict defensive liability in this case but recognize it in others, and his moral status account aims to underwrite this stance.[86] Still, many of the arguments that we will soon consider are aimed, more or less explicitly, not only at rejecting strict defensive liability in this particular case but at rejecting it in a more thoroughgoing fashion.

Before considering these arguments, however, we may observe that this article's first section provides a significant presumptive case for thinking that the threatener in Conscientious Driver *is* liable to defensive killing. For the risk internalization argument about the Sickly Driver cases establishes that at least some acts of entirely permissible and responsible driving do indeed incur strict defensive liability. Moreover, the threatener in Sickly Driver 5 (and its variants) is poised to infringe the very same kind of fact-relative duty against harm as the threatener in Conscientious Driver. Each threatener is about to hit and kill a pedestrian, as a side-effect, with his car. If the risk internalization argument about the Sickly Driver cases is

[86] See the discussion of the moral status account in the previous section of this article, pp. 44 - 50.

sound, therefore, there is a presumption of strict defensive liability for threatening to infringe this sort of fact-relative duty against killing. And this presumption derives additional, more abstract support from the two general moral principles (the presumptive parity principle and risk internalization principle) that we formulated in the course of developing that argument. Both the case-based argument and the general principles support the view that a driver who faultlessly threatens to hit and kill a pedestrian is presumptively liable to be defensively killed.

To be sure, it is possible that this presumption is undercut in cases like Conscientious Driver. Later on in the section, we will examine some reasonably promising strategies for undercutting it. But let us first consider, and reject, several existing philosophical arguments against recognizing strict defensive liability in Conscientious Driver. Our work so far will help to identify where each of these arguments goes awry.[87]

One argument is that, because the conscientious driver has acted in an entirely permissible and reasonable manner - he has done everything we can fairly ask him to do - it would be unfair to deem him liable to be defensively killed.[88] But we have already seen that rejecting strict defensive liability for lack of fairness is misguided.[89] It is indeed unfair, in some morally significant sense, that a person should find himself liable to be defensively killed although he has driven with all due care. But it is similarly unfair that someone should find himself morally required to let himself die, because he finds himself constrained (through no fault or choice of his own) by a stringent fact-relative duty against hitting and killing another person with his car. The stringency of the duty is the ultimate ground of the liability, and the unfairness of

[87] There is one important argument against deeming the conscientious driver liable that I cannot closely examine here. This argument relies on the plausible metaphysical premise that the driver is no more causally responsible than his victim for the harm that he threatens to inflict - metaphysically speaking, the causation of harm is always a joint affair. I believe that the sort of argument developed in this article's first section helps to pinpoint where this line of thought goes awry. What morally distinguishes the driver from his victim is *not* any difference in causal responsibility between them. They are indeed causally responsible, in equal measure, for the harm the driver's victim stands to suffer. Similarly, the tragic situation at hand was equally foreseeable to both parties; they are equally responsible for it in that sense as well. Instead, these parties are morally distinguished by the differing stringency of the fact-relative duties that they are respectively poised to infringe. The stringency of deontological moral duties against causing harm is, in this case and more generally, determined only partly by metaphysical facts about causal relationships to the harm in question. I develop this point, partly as it applies to a parallel debate about liability for accidental harm in tort law, in "The Coherence and Justification of General Strict Liability" (unpublished manuscript).

[88] Ferzan, "Culpable Aggression," p. 681.

[89] See above, pp. 36 – 43.

the former is continuous with the unfairness of the latter. Neither the duty nor the liability should be rejected for lack of fairness unless we are prepared to renounce deontology more generally.

Consider, next, the argument from the *moral gravity* of liability determinations. To say that an agent is liable to some harm is to say that her rights against this harm have been lost (or substantially weakened). But our rights against being killed are among the most significant moral rights we possess. And so, this argument maintains, we should not be able to lose such rights except on the basis of choices that have considerable moral significance.[90] Culpable aggression may do, but driving a car with all due care - or threatening to kill someone when the car spins out of control, through bad luck - will not.

The argument may have some force against the moral responsibility account, in particular, which grounds the conscientious driver's strict defensive liability in the fact that he is somewhat more responsible for his victim's impending death than she is. A slight difference in comparative responsibility for harm may seem too a slender reed on which to hang the conclusion that someone has lost his rights against being killed.[91] The objection has much less force, however, if we ground the driver's strict defensive liability in the stringency of his evidence-relative duty against risking death to the pedestrian. For this duty derives, in turn, from the driver's stringent fact-relative duty against hitting and killing the pedestrian. It is the profound moral significance of this underlying duty against killing that ultimately explains why the driver loses his own moral protection against being killed. By contrast, where a faultless threatener's fact-relative duty against causing lethal harm to another innocent person is much less stringent (as in Unowned Vial, or Threatening Organ Donor),[92] the threatener's choice to risk the infringement of this duty *cannot* render him liable to be killed. This explanatory structure thus provides an intelligible and plausible connection between

[90] Lazar, "Responsibility, Risk, and Killing in Self-Defense," pp. 701; Haque, *Law and Morality at War,* p. 262.
[91] The force of the objection might be blunted by reformulating the moral responsibility account, so that it does not *strip* a faultless threatener of his rights against being harmed, but instead *weakens* those rights (to the point that even an impartial observer may permissibly harm the threatener in order to defend his victim). But the objection would still have some force. Even on this reformulation, the moral responsibility account would still drastically abrogate the strength of a faultless threatener's moral protections on the basis of differences between him and his victim that are (by assumption) comparatively morally insignificant.
[92] See above at pp. 31 – 35.

the loss of a threatener's significant moral protections and the moral significance of the underlying duties whose infringement is at stake.

Next consider the argument from *rights as reasonable demands.* Because the conscientious driver is imposing only very small risks of harm on other people, and doing so justifiably, no one can reasonably demand that he refrain from imposing these risks. Thus no one has a right against his imposing these risks, for rights are intimately connected to (or, perhaps, consist in) reasonable demands.[93] And if that is correct, the conscientious driver cannot be liable to defensive harm, for defensive liability is predicated on the infringement of another person's right.[94] But this argument relies on an untenably cramped conception of the explanatory and deliberative functions that a right against harm might play. One such function is to ground, in appropriate circumstances, a legitimate or reasonable complaint against the behavior which the right proscribes. As we have seen, however, another such function is to guide an agent's own deliberation under uncertainty: to instruct him, for example, that he is obligated to internalize a given risk of harm, if he can, rather than impose this risk on the relevant right-holder. Again, it is ultimately in virtue of an agent's stringent fact-relative duty against inflicting some harm that he has a stringent evidence-relative duty against imposing risks of that harm - such that he is presumptively obligated to internalize such a risk, and liable to defensive measures that will in effect force him to internalize it. That is why, if the underlying fact-relative duty is not very stringent (as in Unowned Vial and Threatening Organ Donor)[95] the agent is *not* obligated to internalize the relevant risk and not liable to defensive harm. Once we recognize that fact-relative and evidence-relative duties (and correlative rights) have this broader moral significance, we should abandon any thought that rights just are, or are co-extensive with, reasonable demands.

Consider, next, the argument that conscientious drivers should not be deemed to incur strict defensive liability, so long as they have driven prudently and in accordance with the rules of the road, because the activity of driving is widely practiced and widely beneficial.[96] On its most natural interpretation,

---

[93] Quong, *Morality of Defensive Force,* p. 154 - 160.
[94] This assumption is itself dubious, since it rules out defensive liability for the infringement of impersonal moral duties.
[95] See above at pp. 31 – 36.
[96] Quong, *The Morality of Defensive Harm,* pp. 27, 35 - 36.

this argument will produce the wrong result in Sickly Driver 5. For Threatener's act of driving himself to the hospital, in that case, is entirely permissible, prudent, and reasonable. His driving is unusually risky, of course, but this need not mean it contravenes any rule of the road.[97] In response, it might be suggested that our specification of the relevant social practice or activity should be more fine-grained: it is the activity of driving in the normal way, so as to impose the normal amount of risk, that is exempt from incurring strict defensive liability. But if we need to gerrymander our specification of the relevant activity in this way, so as to pick out a particular feature of the threatening action in question (namely, the level of risk it imposes), then it would seem that this particular feature of the relevant action is doing critical justificatory work - rather than, or in addition to, its subsumption within any general activity type.

This last observation suggests a different argument that is both simpler and more promising. In Sickly Driver 5, Threatener is imposing a fairly high (1%) risk of death on Victim. The risks that Threatener is imposing in Conscientious Driver are much smaller. This asymmetry might be invoked to explain why Threatener is liable to defensive killing in Sickly Driver 5 but Threatener is not liable in Conscientious Driver, although each is risking the infringement of the same kind of fact-relative duty against harm. Unless it is handled with care, however, this strategy will generate implausible results in other cases. Consider, for example,

> *Tiger Owner:* The inhabitants of a certain small township engage only in activities that pose extremely small risks of harm to one another. They do not drive cars; they do not keep any animals as pets (for there are no suitable animals around); they do not traffic in the transportation of explosives or engage in activities that might cast dangerous fumes into the air; and so on. But one day Threatener, a resident of the township, decides to venture abroad and capture a tiger. After returning home, Threatener patiently domesticates the tiger, skillfully mitigating its wild and violent instincts. In this

[97] If the point is not immediately obvious, it may be established by considering a variant of Sickly Driver 5. Suppose that, some years ago, a different sickly driver in the same town drove himself to the hospital, thereby imposing a 1% risk of death on a different pedestrian. The town council subsequently promulgated special rules of the road to regulate acts of driving which impose unusually large risks of harm. Even if Threatener is complying with these rules, he is liable to defensive killing should the 1% risk of death he is imposing on Victim eventuate.

and other ways, he takes enormous care to ensure that the tiger will never harm anyone in the township. As a result of these precautionary measures, the risk that the tiger will kill someone is now very small - roughly equivalent to the risk of a conscientious driver's well-maintained car going out of control, in a crowded city, and killing a pedestrian. It is therefore evidence-relative permissible for Threatener to keep the tiger, in order to obtain the pleasures of its companionship, just as it is evidence-relative permissible to drive regularly to the cinema. But one day, while Threatener is taking his tiger for a walk in the park, his precautions fail; the tiger is now about to escape from its leash and maul another person, Victim, who is sitting on a bench nearby. Impartial Observer, who knows all the facts about the situation, can save Victim from the tiger only by blowing it up with a rocket launcher before it escapes its leash – but this will unavoidably kill Threatener.[98]

If we exempt the threatener in Conscientious Driver from defensive liability simply on the grounds that he poses too small a risk of harm to incur such liability, then we are committed to exempting the threatener in Tiger Owner from defensive liability on the same basis. But it seems very plausible, intuitively, that this tiger owner *is* liable to be defensively killed. And the risk internalization principle supports this judgment. For imagine that Threatener can install a device on his tiger's collar that will completely the internalize the risks of harm that he is imposing on other people by keeping the tiger and walking it around town. This risk-internalizing collar will monitor the tiger and blow it up, unavoidably killing Threatener in the process, if and only if the collar detects that the tiger is about to escape and kill someone else (*Tiger Owner With Risk-Internalizing Collar*). It seems highly plausible that Threatener is obligated to install this device.[99] And from

[98] The scope of strict tort liability differs between American, English, French and German tort law, but all of these legal systems impose liability for such damage caused by wild animals. See e.g. *Restatement (Second) of Torts* § 509 cmt. d (1977). Such doctrines are insightfully discussed in K. Simons, "The *Restatement (Third) of Torts* and Traditional Strict Liability: Robust Rationales, Slender Doctrines," *Wake Forest Law Review* 44 (2009): 1355 - 1382. For similar cases in the philosophical literature, see J. Coleman, *Risks and Wrongs* (Oxford: Oxford University Press, 1992), p. 257; Tadros, *The Ends of Harm*, p. 231.

[99] This judgment can in turn be supported by the presumptive parity principle. Suppose an agent can save himself from a fate far worse than death either by killing himself or releasing his tiger in the direction of another person (with the certain but unintended consequence that it will kill that person). The agent's fact-relative duty against killing the other person, by releasing the tiger toward him, seems sufficiently stringent that the agent is obligated (in the fact-relative sense) to kill himself instead. If that is correct, the agent will presumptively be obligated (in the evidence-relative sense) to internalize any given risk of death, rather than inflict an equivalent risk of death-by-tiger on the other person

this judgment we may construct a risk internalization argument, in our familiar way, to the conclusion that Threatener is liable to defensive harm if necessary to prevent his tiger from mauling Victim - including in the original case, where he has no such risk-internalizing mechanism available.

So appealing *simply* to the small magnitude of the of risk in Conscientious Driver will not do. We might next try the following hybrid strategy: if an agent is imposing a risk that is sufficiently small, *and* she is engaged in an activity that is widely practiced and widely beneficial, then she is exempt from strict defensive liability. It is not obvious whether a principled rationale for this hybrid strategy can be found, but at least it seems to handle our cases appropriately. There are other cases, however, in which it delivers implausible results. Suppose that the incidence of murderous assault in a certain rural town precipitously increases over the course of a decade. Over the same decade, the capacities of the town's law enforcement apparatus precipitously decline. In these circumstances, the activity of responsibly using protective firearms, in evidently necessary and proportionate self-defense, might become widely practiced and widely beneficial. Nevertheless, if some faultless threatener should attempt to shoot and kill an innocent and unthreatening person - on the basis of her reasonable but mistaken belief that he is trying to murder her - she would be liable to lethal defensive harm on her victim's behalf (*Mistaken Attacker in a Violent Town*). This threatener would be obligated, if she could, to internalize the risk that by shooting at her target she will be killing someone who is in fact innocent and unthreatening. And she would not be relieved of this obligation by the fact that she is engaged in a type of activity - responsibly using a firearm, in evidently necessary and proportionate self-defense - that is widely practiced and widely beneficial.[100] Just as she would be obligated

in question, in order to save himself from a fate worse than death. And he will be similarly obliged to internalize any such risks he creates for the sake of obtaining smaller personal benefits. That is the structure of the situation in Tiger Owner With Risk-Internalizing Collar.

[100] Of course, we could choose to specify the relevant (widely practiced and beneficial) activity type as defensively using a firearm in a way that is *actually* (i.e. fact-relative) necessary and proportionate. The faultless threatener's action would not fall under the umbrella of this activity type. But consistency would then require us to describe the activity type that supposedly renders normally risking conscientious driving immune from defensive liability in fact-relative terms as well. And that will not do, since any conscientious driver whose car threatens to kill a pedestrian has engaged in fact-relative impermissible driving - which activity (so described) is obviously *not* widely beneficial. This particular point suggests a more general concern: in my view, there is no way of specifying the relevant activity types that is both principled and extensionally adequate. For this reason, among others, I do not think our theory of defensive liability should accord any fundamental moral significance to activity types (as opposed to the specific features of specific, threatening actions).

to internalize this risk, if she could, she is liable to a defensive measure that will in effect force her to internalize it.

Consider, finally, the argument that the driver and the pedestrian are *symmetrically responsible* for the latter's impending death.[101] Plausibly, the conscientious driver is no more causally responsible than the pedestrian for this harm: metaphysically speaking, the causation of harm is a joint affair. Plausibly, moreover, it was equally foreseeable to both parties that their respective activities (driving in the vicinity of pedestrians, on the one hand, and walking in the vicinity of cars, on the other) might result in such a tragic situation. Given these symmetries between driver and pedestrian, there is no compelling moral reason for an impartial observer to kill one of these innocent parties in order to save the other. Put another way, there is no morally significant difference between the driver and the pedestrian that can undercut the presumptive moral asymmetry between doing and allowing harm.[102] Or so the argument from symmetrical responsibility maintains.

Defenders of the moral responsibility account might reply by insisting that the driver *is* more causally responsible for the situation at hand, or that the possibility of such a situation was more foreseeable to him than it was to the pedestrian. But it is doubtful that any plausible metaphysical account of causation or defensible conception of foreseeability can underwrite these claims.[103] And it is therefore puzzling, if we assume that strict defensive liability is grounded entirely in causation and foreseeability, why the conscientious driver should be liable to defensive killing. But the puzzle dissolves if we reject this assumption, and instead maintain that a threatener's strict defensive liability is grounded in the stringency of the duties which he is infringing.

---

[101] See Lazar, "Responsibility, Risk, and Killing in Self-Defense." Several decades ago, drawing on the work of the economist Ronald Coase, Stephen Perry developed a very similar argument about strict corrective liability in tort law, in "The Impossibility of General Strict Liability," *Canadian Journal of Law and Jurisprudence* 1 (1988): 147 - 171.

[102] The argument from symmetrical responsibility might be taken to support a general rejection of strict liability. That would be mistaken, however. Suppose that a villain will kill a threatener unless she plays Russian Roulette on an infant victim. The threatener is now about to be shoot and kill the victim. As between threatener and victim, only the threatener's actions are causally responsible for this harm, and the harm is only foreseeable to the threatener. The argument from symmetrical responsibility does not, therefore, block strict defensive liability in such cases.

[103] Perry, "The Impossibility of General Strict Liability."

As we have seen, in our discussion of the Sickly Driver cases, a driver who hits and kills a pedestrian with his car is infringing a highly stringent fact-relative duty towards her. So, too, a driver who risks hitting and killing a pedestrian with his car is infringing a highly stringent evidence-relative duty against imposing risk on her. To drive in this vicinity of pedestrians is, absent some undercutting moral consideration, to infringe a stringent evidence-relative duty of this kind. By contrast, a pedestrian infringes no comparably stringent duty by walking in the vicinity of cars. To bring out the point, consider

> *Sickly Driver and Sickly Runner (Large Risk):* Both Threatener and Victim are suffering from a deadly disease, and each must get to the hospital very soon in order to survive it. Threatener can only get to the hospital by driving there immediately; Victim can only get to the hospital by running there immediately. However, as both parties know, if Threatener drives at the same time as Victim, there is a 99% chance that his car will spin out of control along the way, thereby hitting and killing her. (After this, Threatener will be able to walk himself the rest of the way to the hospital.)

It seems permissible for Victim to run to the hospital, although she will thereby impose on Threatener a stringent evidence-relative duty against driving himself to the hospital - a duty sufficiently stringent that Threatener will then be morally obligated to allow himself to die, unless he has some means (such as a mechanical device) of internalizing the 99% risk of hitting and killing Victim that driving himself to the hospital will create. If this is correct, then Victim's duty against setting back Threatener's interests in this way - by imposing on him a highly stringent duty against driving his car - is not nearly as stringent as the duty against driving that Threatener will thereby acquire. This is true despite the fact that Victim will, by imposing this duty on Threatener, cause him to die (assuming that Threatener complies with this duty, as he will then be obligated to do).

The same basic analysis will hold when the relevant risk of harm is very small, such that the threatener may permissibly impose this risk on the victim even if he cannot internalize it. Consider

> *Sickly Driver and Sickly Runner (Small Risk):* Both Threatener and Victim are suffering from a deadly disease, and each must get to the hospital very soon in order to survive. Threatener can only get to the hospital by driving there immediately; Victim can only get to the hospital by running there

immediately. However, as both parties know, if Threatener drives at the same time as Victim runs, there is a 1% chance that his car will spin out of control along the way, hitting and killing her.

Threatener may permissibly drive to the hospital and Victim may permissibly run there. This is true even if Threatener cannot internalize the 1% risk that he will thereby hit and kill Victim with his car. This risk is sufficiently small that Threatener may, if necessary, impose it on Victim. But if Threatener *can* internalize this risk, it seems that he is obligated to do so (just as in the Sickly Driver cases we discussed in section one). If that is correct, then Threatener's evidence-relative duty against imposing a 1% risk of death on Victim, by risking hitting and killing her with his car, is highly stringent. It is much more stringent than Victim's duty against imposing this duty on Threatener (along with the attendant moral burdens, such as an obligation to internalize the relevant risk and liability to functionally equivalent defensive harm, that this duty brings in its trail).

The basic moral structure of these situations is not anomalous. It is closely akin to the structure of a famous case that we have already mentioned, in which one drowning sailor can outswim another in order to take possession of the single life-saving plank of wood in their proximity.[104] It seems permissible for the first sailor to do this, and once she has done it, it seems impermissible for the second sailor to dispossess her of the plank (by, say, yanking it away from her).[105] So the first sailor's duty against taking possession of the plank, and thereby imposing stringent moral duties on the second sailor, is not nearly as stringent as the duty against dispossession that the second sailor thereby acquires.[106] Similarly, an asthmatic may permissibly use an abandoned respirator in order to save her life, although she will thereby put another

---

[104] See above at p. 35.

[105] The case is often referred to as "The Plank of Carnades." But Cicero attributes it to another philosopher, Hecaton of Rhodes, rather than to Carneades. See Cicero, *On Duties,* eds. M.T. Griffin & E.M. Atkins (Cambridge: Cambridge University Press, 1991), p. 134. Legal theorists discussing this case have tended to focus on its significance for excuse and necessity doctrines in the criminal law. See, for example, K. Ghanayim, "Excused Necessity in Legal Philosophy," 19 *Canadian Journal of Law and Jurisprudence* 31 (2006); J. Hrushcka, "On the History of Justification and Excuse in Cases of Necessity," in *Prescriptive Formality and Normative Rationality in Modern Legal Systems,* eds. W. Krawietz, N. MacCormick, G.H. von Wright (Berlin: Duncker & Humblot, 1994).

[106] Western legal theorists have long debated whether, if the second sailor does choose to dispossess the first sailor of the plank, he should enjoy any excuse for his unjustified action (in light of the great psychological pressure under which he is acting). See Ghanayim, "Excused Necessity"; Hruschka, "On the History." This debate presupposes that the second sailor would act impermissibly by dispossessing the first sailor of the plank. It also implicitly presupposes that the first sailor commits no such serious wrong against the second by taking initial possession of the plank.

asthmatic under a stringent moral duty against dispossessing her of the respirator - and, in this way, foreseeably cause his death, supposing that he will honor his resulting duty against dispossession (or supposing that it will be enforced, say by an impartial observer). Or imagine, to provide a final illustration of the same general phenomenon, that two people (Alain and Beatrice) are trapped on an island that will soon be destroyed by a volcanic explosion. There is a single bridge off the island, and it can support only a single person's weight; moreover, as soon as one person crosses the bridge, it will collapse into the chasm below, leaving the other person stranded on the island. It seems permissible for Alain to outrun Beatrice in order to cross the bridge and get off the island. And it seems impermissible, once he has done so, for Beatrice to shake him off the bridge in order to use it herself. So Beatrice's duty against killing Alain, in this way, is much more stringent than Alain's duty against imposing such a duty upon her.

Intuitive judgments of this kind indicate that an agent's fact-relative and evidence-relative duties against imposing stringent fact-relative and evidence-relative duties on another person will often possess only limited stringency. These intuitive judgments are not beyond doubt or immune to revision. In particular, they sit in considerable tension with inherently plausible ideas about egalitarian fairness under conditions of material scarcity.[107] But this tension seems to me continuous with the kinds of tension between deontological morality and egalitarian fairness that we have already encountered.[108] In my view, it is preferable to endorse the relevant deontological judgments than to revise ordinary moral thought in the drastic ways that abandoning these judgments would require.

But my concern here is not to argue for this methodological stance about this class of cases.[109] It is to illuminate the structure of cases like Conscientious Driver, by showing that they fall within this broader class. In all of these cases, a threatener has a stringent duty against (risking) harming a victim, and he has this duty only because the victim has chosen to act in a certain way (by walking alongside a road, or taking

[107] I explore this tension at greater length in "Property Rights in Dire Straits" (unpublished manuscript).
[108] See above, pp. 36 – 43.
[109] In "Property Rights in Dire Straits," I argue that rejecting these kinds of judgments will likely require us to reject deontological moral judgments that are even more deeply rooted, such as judgments about the impermissibility of appropriating artificial or even natural body parts. Even so, I am not at all sure that any of these judgments are defensible on reflection, which is one reason among many that I am an uneasy deontologist.

possession of a plank, etc.). In so acting, the victim may also infringe a duty towards the threatener. But this duty is much less stringent than the duty which the threatener thereby acquires. And so the threatener is liable to severe defensive harm, in these cases, although the harm he is about to inflict on the victim will be the equally foreseeable, joint causal product of their actions. For it is the stringency of a faultless threatener's duty against risking harm to another person that explains his liability to defensive harm. This analysis applies where, as in the Sickly Driver and Sickly Runner cases, the relevant risk of harm is significant and the parties' reasons for acting are morally weighty. But it also applies where, as in Conscientious Driver, the relevant risk of harm is small and the parties' reasons for acting are morally insignificant.

We have considered and rejected several arguments against recognizing strict defensive liability in Conscientious Driver. It may now seem, therefore, that we should simply accept that the faultless threatener in this case is indeed liable to defensive killing. That would be an entirely reasonable conclusion to draw, in my view, and I do not think we should bend backwards in order to resist it. Nevertheless, this conclusion should not be drawn too hastily. For even if there is a presumption of strict defensive liability for car driving, it is possible that this presumption is undercut in Conscientious Driver (or, more precisely, certain variants of it). There are two strategies that seem to me reasonably promising. The first strategy is not infrequently mentioned by philosophers writing on strict defensive liability,[110] but it has rarely been discussed or developed at much length. The second is, to my knowledge, more novel. Both strategies are compatible with recognizing a presumption of strict defensive liability for harms caused by car driving: both seek to undercut this presumption, in a principled way, in certain cases.

FIRST STRATEGY: RECIPROCAL RISK IMPOSITION. It is easy enough to imagine that, in Conscientious Driver, Threatener is imposing a risk of death on Victim that is no greater than the risks of death that Victim has previously imposed on him. For it is easy enough to imagine that Victim has previously driven a car around town herself (*Conscientious Driver Threatening Car-Owning Pedestrian*). If we carefully specify the details of the case in this way, so as to make it clear that Victim has previously imposed

[110] See e.g. Ferzan, "Culpable Aggression," p. 678; McMahan, "Basis of Moral Liability," p. 404.

an equivalent risk of death on Threatener, it seems plausible that Threatener will be relieved of his presumptive obligation to internalize the risk that he is imposing on Victim. And if that is the case, then Threatener will not be liable to lethal defensive measures that will in effect force him to internalize this risk.

A similar analysis may hold if Victim is *simultaneously* imposing an equivalent risk of harm on Threatener. For suppose that Victim is not a pedestrian walking by the side of the road; she is a driver, proceeding in her own car to the cinema. Threatener's car is now about to spin out of control and crash into Victim's car, killing her (*Conscientious Driver Threatening Another Conscientious Driver*).[111] It seems intuitively plausible to me that Threatener is not liable to be defensively killed in this case. Nor, I think, would he be obligated to internalize the risk of harm he is imposing on her, even if he could. Finally, it may also be plausible (if somewhat less so) that Threatener is not defensively liable if Victim has previously imposed equivalent risks on *other* people, but not on Threatener. This might be true if, say, Threatener has just moved to town (*Newly Arrived Conscientious Driver Threatening Car-Owning Pedestrian*). In all of these cases, I will say - somewhat unnaturally, but this does not matter for now - that Threatener is imposing a *reciprocal risk* on Victim.[112]

By contrast, if we suppose that Victim is not imposing, and has never imposed, any remotely comparable risk of harm on Threatener or anyone else, the moral situation seems to differ considerably. Imagine that Victim is a committed environmentalist who never puts a car on the road, and whose other activities are extremely innocuous. Now Threatener is driving himself to the cinema to see the new movie premiere, and Victim is walking there; Threatener's car spins out of control and will kill Victim unless an impartial observer blows it up (*Conscientious Driver Threatening Environmentalist Pedestrian*). It seems plausible to me that Threatener is liable to defensive harm in this case - certainly, much more plausible than in Conscientious Driver Threatening Car-Owning Pedestrian.

---

[111] Compare the cases in Ferzan, "Culpable Aggression," p. 678; McMahan, "Basis of Moral Liability," p. 404.
[112] The label does not naturally describe the last case, in which Threatener is imposing a risk on Victim, who has in turn imposed an equivalent risk on someone *else.* See below, p. 66.

This judgment can be supported by the risk internalization principle. For suppose that Threatener can install on his car a mechanical device that will completely internalize the small risk that, by driving himself to the cinema, he will hit and kill Victim: the device will continuously monitor the car's surroundings and blow it up if and only if it is about to kill Victim. Nor will the device have any other relevant effects; it will not, for example, internalize the risks of harm that Threatener is imposing on other *drivers,* or (for that matter) other pedestrians. It seems plausible that Threatener must install this risk-internalizing device. If he chooses to drive himself to the cinema, and thereby risk the death of a civilian who has never done the same to him or anyone else, it seems difficult to maintain that he may offload onto her the risk of death that his self-benefiting action entails. He would be similarly obligated to request a functionally equivalent course of defensive monitoring from an impartial observer. If this is correct, then the risk internalization principle will imply that Threatener is liable to be defensively killed in order to prevent him from killing the environmentalist pedestrian, even if he has no such device available.

Our intuitive judgments about these various threateners' obligations to internalize risk (if they can) thus seem to vary alongside our intuitive judgments about their liability to defensive harm. Both sets of intuitions accord considerable significance to the reciprocal or non-reciprocal character of the risk a threatener is imposing. This suggests a promising general account of all these variants of Conscientious Driver. In each case, we say, the driver is *presumptively* obligated to internalize the risk of hitting and killing another person with his car (just like the drivers in Choice of Lethal Harms, Choice of Large Lethal Risks, Choice of Small Lethal Risks, and Sickly Driver 3). But this presumption is undercut if and only if the victim has previously imposed, or is currently imposing, a comparable risk on the driver (or, perhaps, on someone else). The risk internalization principle will then imply that Threatener is liable to defensive killing, in order to prevent his car from killing his victim, if and only if the victim is not imposing and has not previously imposed any such risk on him (or, perhaps, on someone else).

This strategy is appealing, in part, because there are independently compelling grounds for attributing this sort of undercutting moral significance to reciprocal risk imposition (as we saw in our

discussion of Reciprocal Russian Roulette).[113] But, of course, the mechanics of the reciprocal risk view must ultimately be worked out and defended in greater detail. What makes two risks of harm stand in the relevant relationship of reciprocity, such that the threateners who impose them are not obligated to internalize these risks or defensively liable for these risks' eventuation? Must the risks be strictly equivalent, or can one be slightly larger than the other? If so, by how much? How should the view handle cases where A imposes a risk on B who in turn imposes a risk on C? Where C in turn imposes a risk on A? If these are not cases of reciprocal risk imposition, strictly speaking, should they nevertheless be treated in the same way?

As I will discuss in the sequel to this article, considerable difficulties arise in attempting to answer such questions.[114] Because I cannot opine about the considered defensibility of the reciprocal risk account - as we might call this general approach - until that discussion, I cannot yet say whether it will successfully relieve certain conscientious drivers of strict defensive liability. For now I must simply observe that the reciprocal risk account promises to yield intuitively plausible verdicts about this domain. This gives us reason to carefully develop and assess the reciprocal risk account, but it does not foreordain that the account will prove defensible when fully developed.

Even if the non-reciprocal risk account ultimately proves to be sound, some philosophers may continue to resist the idea that an agent can incur strict defensively liability merely by engaging in conscientious, normally risky driving. These philosophers might insist that *even if* a given conscientious driver is imposing an obviously non-reciprocal risk (as in Conscientious Driver Threatening Environmentalist Pedestrian), he is exempt from defensive liability should this risk materialize. This intuition is not so ironclad, in my view, that it should be taken to refute an otherwise plausible theory of strict defensive liability. Nevertheless, at least in a certain frame of mind, I can feel its force. So it is worth considering how this position might be explained and defended, consistently with the presumptive parity principle and the risk internalization principle. Here is a speculative strategy for doing so.

---

[113] See above at p. 42.
[114] "Strict Liability to Defensive Harm II," pp. 109 – 125.

SECOND STRATEGY: NORMATIVE POWERS OVER PUBLIC PROPERTY. Suppose my backyard contains a large body of water which I freeze over into an ice rink for the winter. I invite all my neighbors to skate on it, but I announce to them that, if they choose to do so, they will absolve me of any defensive or corrective moral liability for harms they might suffer in consequence: "skate at your own risk."[115] It seems plausible that, within limits, I possess a normative power to do this: by suitably communicating such an announcement, I can release myself from certain moral liabilities that I would otherwise incur.[116] And this is not simply because, if a neighbor skates on my rink, he signals consent to my terms (in any non-fictive sense). After all, I would seem to be absolved of liability with respect to a given neighbor even if, before getting onto my rink, he shouts out "I do not accept your terms!" The point is instead that my moral ownership of the pool generates or constitutively involves an entitlement to determine the terms on which others may access and use it. And I might stipulate, as one such term of use, that people who use my pool thereby waive my prospective moral liabilities for harms they suffer in consequence.

Arguably, however, I also possess the moral power to waive (through suitable communication) the default moral liabilities of *other* people who use my rink. I might say, for example: "I hereby announce that anyone is free to access this ice rink on condition that she absolves all other people who threaten to cause her harm, by using the rink in a conscientious and permissible way, from any form of liability for doing so." Arguably, by making this announcement, I can release people who use my ice rink from defensive and corrective liability for resulting harms (including harms that result from non-reciprocal risk impositions, as when some skaters do ice tricks, or skate especially quickly).[117] Similarly, I might choose to release people who use my rink from their presumptive risk internalization obligations, by announcing: "I hereby announce that anyone is free to access this ice rink on condition that she releases all other people who might cause her harm from their ordinary obligations to internalize the risk of doing so."

---

[115] Cf. *Smollett v. Skating Dev. Corp.* 793 F.2d 547 (3d Cir. 1986).

[116] It is much less obvious that I can absolve myself, by prospectively making such a communication, of liability for harms that I *culpably* cause to people who use my ice rink.

[117] This is most plausible with respect to others who hear and understand the announcement, and somewhat less clear if and when someone does not. Cases of this kind thus implicate murky questions regarding the kind of communicative uptake that is necessary for the successful exercise of a normative power.

The existence of such normative powers is consistent with the presumptive parity principle and the risk internalization principle. Together these principles imply that, if an agent's fact-relative duty against harming another person is sufficiently stringent, then he is presumptively obligated to internalize any given risk of the relevant harm, and presumptively liable to defensive measures that will in effect force him to internalize this risk. The sort of normative power we are now considering is a mechanism for undercutting, within a limited domain, either or both of these presumptions.[118]

If this is correct, an instance of the same general phenomenon might be at work - more murkily and more contestably - in Conscientious Driver. For roads and adjoining walkways are typically understood as public goods, which belong to the state or the community at large.[119] And if the state or community morally owns these public spaces, it may have the power to modify the default moral rights and liabilities of people who choose to use these spaces, just as I have the power to alter the default moral rights and liabilities of people who choose to use the ice rink in my backyard.

The question then arises, of course, whether any given state or community has actually exercised such a moral power over the use of its public spaces. In our own society, the answer may well be yes. After all, the law of torts is familiar with accidental injury inflicted by faultless driving. And in Anglo-American jurisdictions, at least, the law conspicuously declines to impose tort liability for such injury.[120] Indeed, even the leading legal defenders of strict tort liability have sometimes made conspicuous exception for injuries suffered while using public roads.[121] That said, it is not obvious how much evidentiary significance such

---

[118] Importantly, I might choose to exercise only my normative power to relieve ice rink users of their obligations of risk internalization without similarly exercising my normative power to relieve them of defensive and corrective liability (or vice versa). This is yet another illustration of the general phenomenon that, although the obligation to internalize a risk and defensive liability for the eventuation of that risk will *presumptively* go together, this connection can be severed by undercutting factors.

[119] The pioneering discussions of roads in contemporary political philosophy is in A. Ripstein, *Force and Freedom: Kant's Legal and Political Philosophy* (Cambridge: Harvard University Press, 2009): 232 - 266.

[120] See generally *The Development of Traffic Liability,* ed. W. Ernst (Cambridge: Cambridge University Press, 2014). For an appropriately critical overview of the broader legal treatment of car driving, see G.H. Shill, "Should Law Subsidize Driving?" *New York University Law Review* 95 (2020): 498 - 579.

[121] The English judge who most famously and eloquently defended a default rule of strict tort liability was Baron Bramwell. An example:

legal materials should be accorded. The matter is substantially complicated by the widespread operation of liability insurance, administrative mechanisms for the routinized settlement of tort claims, and the like.[122] These mechanisms substantially blunt the ostensible effects of liability (or no liability) determinations, and thereby obscure the conventional social judgments these determinations can be said to express.

Even if we leave the legal materials aside, however, it seems relevant that our informal social norms make light of the decision to drive with ordinary prudence (in such a way that imposes only very small risks of injury on others).[123] The quotidian decision to drive is not perceived as particularly morally significant, even when undertaken for the sake of acquiring comparatively trivial personal benefits (such as seeing a movie at the cinema). The moral situation is perceived much differently, both by the law and social custom, when someone faultlessly injures another by walking around a tiger, or transporting explosives - even when she acts with the very same level of care, and imposes no greater an amount of risk, and seeks to secure no less weighty a personal benefit.[124] Imposing risk by engaging in such activities is conventionally perceived, it seems to me, as a morally grave affair. So too is imposing risk by driving a car in an unusually risky way, even for excellent reason (as in the Sickly Driver cases).

This constellation of legal positions and social understandings may well rest on underlying moral sentiments that are arbitrary or indefensible. But the moral owner of private property may sometimes enjoy (within limits) the power to determine the terms on which people access and use that property, *even on the*

> "It is just and reasonable that if a person uses a dangerous machine, he should pay for the damage which it occasions; if the reward which he gains for the use of the machine will not pay for the damage, it is mischievous to the public and ought to be suppressed, for the loss ought not to be borne by the community or the injured person. If the use of the machine is profitable, the owner ought to pay compensation for the damage." *Powell v. Fall* 5 Q.Victoria. at 600 - 601.

But here is Bramwell in another case, declining to recognize a right to compensation for injury suffered while using a public road:

> "For the convenience of mankind in carrying on the affairs of life, people as they go along the roads must expect, or put up with, such mischief as reasonable care on the part of others cannot avoid." *Bamford v. Turnley*, 10 Ex. 261 at 267.

[122] See e.g. S. Issacharoff and J.F. Witt, "The Inevitability of Aggregate Settlement: An Institutional Account of American Tort Law," *Vanderbilt Law Review* 57 (2004): 1571 - 1636.

[123] Moral philosophers have occasionally observed and decried this situation. See e.g. D. Husak, "Vehicles and Crashes: Why is this Moral Issue Overlooked?" *Social Theory and Practice* 30 (2004) 2: 351 - 370.

[124] These activities typically attract strict tort liability in both common law and civil law jurisdictions.

basis of her mistaken moral convictions. In the same way, the state or community might enjoy the moral power to determine - on the basis of social judgments that embody moral mistakes - the terms on which people will access and use its public spaces. To be clear, I am not suggesting that the state or community has any *general* moral power to modify our ordinary defensive and corrective moral liabilities. In my view, the law cannot extinguish or substantially modify an agent's defensive and corrective moral liabilities if, in order to avert great peril to herself, she breaks into another innocent person's home and breaks his nose. But if the state or community morally owns public spaces like roads and parks, perhaps it does have the more limited power to alter the moral liabilities we would otherwise incur in the course of using these public goods. I have suggested it is somewhat plausible, if contestable, that (many) Anglo-American jurisdictions *have* made such a determination with respect to ordinary driving on roads and (certain) other public spaces. If it is correct, this suggestion would explain the intuition that the threatener in Conscientious Driver Threatening Environmentalist Pedestrian does not incur strict defensive liability - although he is threatening to kill her by imposing on her a non-reciprocal risk of death.

If these claims are plausible, the case for strict defensive liability for accidental harm will tend to be particularly compelling when (a) the faultless threatener at issue is imposing a non-reciprocal risk on his victim and (b) no societal disposition of a public space, of the kind we have just discussed, is at issue. Both are the case in Tiger Owner. Both are also true in certain cases involving normally risky car driving. Suppose, for example, that a conscientious driver's car goes out of control and threatens to crash into the environmentalist pedestrian's home, killing her (*Conscientious Driver Threatening Environmentalist Pedestrian In Her Own Home*).[125] Intuitively, it seems quite plausible to me that this driver is liable to be defensively killed. And there is not, I suspect, any convincing strategy for undercutting the presumption of strict defensive liability in this case. Neither appealing to reciprocal risk imposition, nor positing that a society can alter the ordinary moral rights and liabilities of people who choose to use the spaces it owns, will do so - unless, perhaps, the state or community is the ultimate moral owner of all the property within its jurisdiction.[126] If that is so, the

[125] A similar example is offered in J. Coleman, *Risks and Wrongs,* p. 256.
[126] S. Nili, "The Idea of Public Property," *Ethics* (2020).

state might enjoy a capacious moral power to alter the ordinary rights and liabilities of people within its jurisdiction, even when they are using their (ostensibly) private property.[127]

If these remarks are plausible, the claim that widely practiced and widely beneficial activities do not incur strict defensive liability may be a partly accurate moral generalization, one that holds within a limited domain. If an activity is widespread, engaging in it (in the normal way) will often impose only reciprocal risks of harm on others. And if the activity is not only widespread but perceived to be widely beneficial, it may be especially likely that the community has immunized it from incurring strict defensive liability (when it issues in injury to other people using public property). But these things will not always be true, and the relevant generalization will thus have important exceptions, even within its limited domain. Here is a somewhat fanciful illustration of the point:[128]

> *Widespread Tiger Walking:* One fine day, most of the residents of a certain small township simultaneously and separately decide to acquire tigers, properly domesticated and trained so as to impose only very small risks of harm on other people (no greater than the risks typically imposed by car driving). Over the following week, these residents proceed to walk their tigers in the public square, thereby imposing small risks of harm on other people using it. Now one such tiger threatens

[127] Even if that capacious view of public property were accepted, moreover, it would not immunize all instances of normally risky driving from strict defensive liability. Suppose, for example, that two small countries - Urbania and Ruritania - are located in close proximity to one another. One decade ago, Urbania went through a rapid spurt of economic development and industrial modernization. Thus its inhabitants now drive cars to the cinema, transport potentially hazardous chemicals in trucks, and engage in other activities that create the kinds of small but non-trivial risks of harm characteristic of life in modern industrial societies. Ruritania has enjoyed no such spurt of economic development (although it has indirectly received many benefits, such as access to cheaper medicines, from the economic development of its neighbor). Thus Ruritanians generally impose much smaller risks on Urbanians than Urbanians impose on them. One day, Urbanian resident Threatener is driving his car to the cinema when it suffers a freak mechanical failure, and careens over the Ruritania/Urbania border, toward a Ruritanian Victim who is reading in his local park. The car will now hit and kill Victim unless impartial observer blows it up, killing Threatener. It seems intuitively plausible to me that Threatener is liable to be defensively killed in this case - more plausible than in the original Conscientious Driver case, and more plausible, even, than in Conscientious Driver Threatening Environmentalist Pedestrian. This makes sense, on the framework I have proposed. Threatening to kill someone by driving a car presumptively incurs strict defensive liability, unless some further morally relevant factor undercut this presumption. In this case, neither reciprocal risk imposition, nor a community's power to modify the default moral rights of people who suffer injury while using its property, can do the necessary undercutting work. And this is true even if we maintain that a state or community morally owns all the property within its jurisdiction.

[128] The case is inspired by the medieval jurist Jacques de Revigny's discussion of the relationship between instantaneous social practice and customary law. See L. Waelkens, *La Theorie De La Coutume Chez Jacques de Revigny* 485 (*repetitio ad Dig.* 1.3.32, § 2) (1984).

to maul and kill a person who is sitting and reading on a bench in the public square. An impartial observer can save this person only by blowing up the tiger, which will kill its owner.

Given enough time, this community might implicitly or explicitly decide that engaging in tiger walking, in (certain) public spaces, should not incur its default moral liabilities. And the community might implicitly or explicitly communicate to its members that such a determination has been made. But no such determination is effected *simply* by the proliferation of this behavior in the community. And before any such determination is made, I believe, walking one's tiger in a public space - and thereby threatening to kill another person using this space - will incur strict defensive liability. This will be true even if the recent uptick in tiger walking is widely beneficial, because, say, it is attracting an influx of economically beneficial tourism.

In my view, therefore, the critical moral dynamics of cases like Conscientious Driver exhibit much greater variety than either defenders or critics of the moral responsibility account have tended to suppose. Structural ideas such as the risk internalization principle and the presumptive parity principle do not settle the moral dynamics of any such case, but they do productively guide our thinking about them. Most importantly, these principles suggest a presumption of strict defensive liability for harms caused by driving - a presumption vividly exhibited and reinforced by the Sickly Driver cases. Philosophers who reject this presumption are under some pressure to explain just how, in their view, cases of unusually risky but entirely permissible car driving (such as the Sickly Driver cases) ought to be analyzed. I doubt that can be done in a way that is both intuitively plausible and attractive on reflection.

Certain arguments about how this presumption might be undercut (such as reciprocal risk imposition and societal dispositions of public property) are reasonably promising, in my view. Others (such as the fact, taken by itself, that the activity of driving is widely practiced and widely beneficial) are less promising. It may be the case that "rights are adjusted to that which happens frequently,"[129] but only when and because there is some specific explanation, more restricted and tailored in character, of why this is so.

My appraisal of these particular arguments for undercutting strict defensive liability in

[129] Odofredus, *Lectura Super Digesto Veteri* 16r. (reprint 1967) (1550) (*repetitio ad Dig.* 1.3.32, § 15).

Conscientious Driver and its variants might of course be contested. But I have been equally concerned to illustrate the shape that arguments of this kind will, in my view, properly assume.

V. Strict Defensive Liability for Altruistic Wrongs

My focus until now has been on strict defensive liability for entirely self-interested wrongs. I have been concerned, in other words, with the liability of faultless threateners acting entirely for their own benefit. Suitably extended, however, the analytical framework I have been developing can illuminate whether and when a faultless threatener acting entirely or partly for the benefit of other people will become liable to defensive harm. My aim here is not to develop this extension comprehensively, but to suggest its general form and exhibit some of its attractions. The extension is incomplete, in large part, because it relies on intuitive claims about the stringency of various altruistic threateners' fact-relative duties against harm and evidence-relative duties against risking harm. For reasons I will soon explain, I have no general account of this matter to offer (in the way the presumptive parity principle offers some general guidance on the stringency of entirely self-interested threateners' duties).

We have already seen one example, Altruistic Injection, in which a faultless threatener is acting for the benefit of another person. In that case, the threatener is acting in order to benefit the very person whom his action now threatens to harm. But philosophers have mostly focused on another sort of case, in which a faultless threatener is acting for the benefit of some people and threatening harm to others. Here is the most discussed sort of example:

> *Fighter Pilot:* A retired commercial pilot decides to fly a single military mission on behalf of a foreign guerrilla army, which is fighting a just war against a villainous dictator. The aim of the mission is to bomb and destroy the dictator's munitions factory, thereby preventing the dictator from killing twenty innocent people. By flying this mission, the pilot will incur a significant risk of death, so it would be permissible for her to refrain from flying it. But she has decided, heroically, to fly the mission anyway. Unfortunately, the pilot's bomb now threatens to kill a civilian bystander. An

impartial observer can save the civilian bystander only by shooting down the pilot, killing her (this will not hinder the pilot's mission, which will succeed in either case).[130]

It seems to me, like many philosophers (though not all), that this sort of threatener is not liable to be defensively killed. This seems true although she is about to infringe a fact-relative duty against killing another innocent person - a sort of duty that is, ordinarily, highly stringent. And it seems true even if the pilot's action is fact-relative impermissible, because (say) her evidence that destroying the factory will prevent the dictator from murdering the twenty innocents is in fact inaccurate.

This view is not uncontroversial, but for now let us suppose it is correct, and see how it might be explained and defended. Perhaps the most influential explanation has been offered by Jeff McMahan.[131] He suggests that that, when a faultless threatener enjoys sufficient other-regarding, evidence-relative moral justification for the threat of harm that she is posing, she will not be liable to defensive harm. In short, *justification defeats liability* (JDL). Since the pilot enjoys such a justification for her action, JDL underwrites the intuition that she is not liable to defensive harm.

But JDL produces counterintuitive results in other cases. Suppose that a villain threats to kill five innocent people unless one of them, Threatener, plays Russian roulette on a passing stranger, Victim. Threatener has sufficient other-regarding moral justification for risking harm to the stranger. Nevertheless, it seems to me, he is liable to lethal defensive harm should the risk he is imposing on the stranger eventuate. The risk internalization principle supports this view. Plausibly, Threatener would be obligated to play Russian roulette on himself rather than the stranger, if he could; he would be obligated to internalize, in this way, the risk of death he is creating partly for his own benefit.[132] Similarly, Threatener would be obligated to internalize this risk by asking an impartial observer to defensively monitor him and kill him, if

[130] My version of this case stipulates that this pilot does not hold any uniformed military role or other institutional office, in order to bypass the complication that the duties of people who hold such offices may be especially stringent, in virtue of their promissory commitments or considerations of role morality. See J. McMahan, "The Just Distribution of Harm Between Combatants and Noncombatants," *Philosophy & Public Affairs* 38 (2010) 4: 343 - 379. McMahan, "Self-Defense Against Justified Threateners," insightfully distinguishes and discusses several variants of this sort of case.
[131] McMahan, "Self-Defense Against Justified Threateners."
[132] This judgment can, in turn, be supported by the presumptive parity principle. For suppose that a villain will inflict a fate much worse than death on Abe and five other people unless Abe shoots and kills a passing stranger, or shoots and kills himself. It seems plausible that Abe must shoot and kill himself rather than the stranger.

necessary, in order to save his victim. And so I think that Threatener is liable to such a defensive measure, even if he cannot actually request it (or otherwise internalize the risk that he is imposing on Victim). Such a defensive measure in effect forces him to internalize a risk that he would be obligated to internalize, if he could.

Perhaps JDL might reformulated, so as to maintain that only *entirely* other-regarding justification defeats liability. Even so, I think, it would have implausible results. Suppose that a villain will kill Threatener's five children unless he shoots himself in the foot or shoots the foot of a passing stranger Victim. It seems plausible that Threatener is fact-relative obligated to do the former. He may not offload the cost of saving his children onto a stranger when he can shoulder this risk himself. Next suppose that the villain credibly threatens to kill Threatener's five children unless he shoots a revolver with one loaded chamber at his foot, imposing a one-in-six chance of a mangled foot on himself, or shoots the revolver at the foot of passing stranger Victim, imposing a one-in-six chance of a mangled foot on her. It seems, again, that he must do the former.[133] And so, I believe, even if Threatener's only option for saving his children is to risk imposing a mangled foot on the stranger, he is liable to defensive harm that will in effect force him to internalize this risk. Thus an impartial observer may defensively shoot and mangle Threatener's foot in order to prevent Threatener from shooting and mangling Victim's foot. The risk internalization principle implies that this threatener is liable to defensive harm, while JDL implies that he is not. In my view the former implication is more plausible.

We might still accept JDL, despite these implications, if doing so were necessary to explain the intuition that strict defensive liability does not attach in cases like Fighter Pilot. But another explanation of this intuition is available to us. Our starting point is the intuitively plausible assumption that the pilot's fact-relative duty against killing the civilian bystander, as a means of destroying the factory, has only limited stringency. To see this, imagine that it were possible for the pilot to destroy the factory by flying a suicide

[133] The same analysis holds, I believe, if the relevant harm is *death*, rather than a mangled foot. But in that case, there is the added complication that, if he kills himself, Threatener will be unable to discharge (some of) his special duties toward his children going forward. That factor *might* make it permissible to instead risk death to the stranger. No such complication is likely to be involved if the relevant risk is a mangled foot.

mission into it, rather than by dropping the bomb (thereby killing the civilian). It seems to me that the pilot would be (fact-relative) permitted do the latter: her duty against killing the civilian bystander, in order to achieve her just aim, is not so stringent that she must achieve this just aim by killing herself instead.

If this is correct, then the stringency of the pilot's evidence-relative duties against imposing *risks* of death on the civilian bystander (in the course of attacking the factory) will be similarly limited. Thus, for example, it would be evidence-relative permissible for the pilot to attack the factory by imposing a 50% risk of death on the civilian, rather than by imposing a 50% risk of death on herself. But suppose that the pilot can *only* attack the factory by imposing a 50% chance of death on the civilian, and is about to do so. Suppose she will thereby kill the civilian, and an impartial observer knows this. The risk internalization principle implies that the pilot is not liable to be defensively killed: the impartial observer may not, in effect, force her to internalize a risk that she would not be obligated to internalize even if she could.

This conclusion seems plausible to me. Such a threatener - who goes beyond the call of duty entirely in order to benefit others, imposes only proportionate harm as a means of doing so, and acts with all due care and moral concern - does not plausibly forfeit her ordinary moral rights against being killed. It is worth noting, however, that some philosophers reject this position.[134] Adam Hosein, for example, argues that the fighter pilot *is* liable to be killed. He defends a Principle of Just Beneficence, according to which "if you are going to be beneficent you must do the beneficent action that best conforms to what justice requires."[135] Presented schematically: "If C harms B she will wrong him, whereas if she harms herself no-one will be wronged. It is more just to do the action that involves less wronging. Thus, if C is going to save A she must do so by harming herself not by harming Victoria."[136] Adil Haque similarly suggests, with an approving citation to Hosein's discussion, that "Whenever possible, we should bear the costs of our own permissible

[134] See, for example, Haque, *Law and Morality at War;* Hosein, "Are Justified Aggressors a Threat"; Rodin, "Justifying Harm," pp. 94 - 95;
[135] Ibid. p. 95.
[136] Actually, the quote is referring to the characters in *Machine,* a hypothetical case that Hosein offers. But he intends for these characters to illustrate a broader moral principle, his Principle of Just Beneficence. I should mention that Hosein arrives at this principle by means of a casuistical argument regarding Machine that is similar to the risk internalization arguments I have offered in this article (although it appeals to case-based judgments about duties to internalize harm, and does not discuss duties to internalize risk). As I argue in the text above, however, Hosein does not (in my view) correctly formulate the structural principles that explain the force of such casuistical arguments.

choices, as well as of discharging our own duties."[137] Thus Haque concurs with Hosein's judgment that the pilot is liable to be defensively killed.

But such principles cannot possibly be correct. Suppose an agent can save five people from a bomb by jumping on it, killing herself, or by kicking it in the direction of the bystander, blowing off his foot. It seems plainly permissible for the agent to do the latter. But by doing so, she would wrong the bystander; she would infringe a *pro tanto* duty against harm that she owes to him. If she instead throws herself on the bomb, she will wrong no one. Hosein's Principle of Just Beneficence thus seems to imply that the agent must kill herself, in order to save the five, if she chooses to the save the five at all. Haque's suggestion that we should bear the costs of our permissible choices whenever possible has the same implication. But this is a deeply implausible result. The agent's *pro tanto* duty against taking off the bystander's foot is not so stringent that he must kill himself, rather than infringe this duty, as a means of supererogatorily saving the five.

There is better way to defend the conclusion that the fighter pilot is liable to defensive killing - if that is the conclusion one wishes to reach. I have suggested that the pilot's fact-relative duty against killing the civilian, in the course of destroying the factory, is not terribly stringent: she would not be fact-relative obligated to destroy the factory by killing herself instead. This premise seems intuitive to me, and it can be further supported by additional philosophical argument.[138] But it is controversial, and some philosophers explicitly reject it. Haque is one of them. He suggests that the pilot *would* be obligated, if she could, to destroy the factory by flying her plane into it on a suicide mission, rather than by dropping a bomb and killing the

[137] Haque, *Law and Morality at War,* p. 79.

[138] Here is the argument. Suppose, for *reductio*, that the pilot's duty against killing the one is sufficiently stringent that she must destroy the munitions factory by killing herself rather than by killing the one. And suppose the pilot is a moral savant, who has correctly ascertained this moral truth. But still, the pilot does not wish to sacrifice herself in order to save the twenty innocents whom the dictator threatens. And it is entirely permissible for her to refrain from making this sacrifice. Then the pilot will, after ascertaining these moral truths, simply refrain from destroying the munitions factory. She will thereby allow twenty innocent people to die - instead of saving these twenty people by killing the civilian, an action that would have been independently proportionate and permissible if she had been unable to sacrifice herself for this cause instead. It seems bizarre that morality should generate such perverse incentives (incentives not to perform independently proportionate, permissible, and impersonally optimistic actions) for immaculately conscientious moral geniuses who, despite being willing to shoulder risks that are considerably beyond the call of duty, retain an entirely permissible degree of self-interest. Even if such a moral system is not internally inconsistent, strictly speaking, there is some broader sense in which its coherence is open to doubt.

civilian.[139] This is so, he believes, despite the fact that pilot would be permitted to refrain from flying the mission at all. In Haque's view, therefore, the pilot's fact-relative duty against killing the civilian, for the sake of destroying the factory, is a highly stringent one. If this correct, then the pilot's evidence-relative duty against imposing any given risk of death on the civilian will also be highly stringent: the pilot would be obligated, if she could, attack the factory by imposing this risk on herself, rather than by imposing it on the civilian. And then the risk internalization principle will imply that, even if the pilot cannot internalize the risk she is imposing on the civilian, she is liable to defensive harm that will in effect force her to do so.

This is, I believe, the most plausible way to defend the conclusion that the pilot is liable to defensive harm: by arguing that the pilot's fact-relative and evidence-relative duties against (risking) killing the civilian *are* highly stringent.[140] I disagree with this claim, of course. But it is beyond the scope of this article to settle this disagreement. Doing so would require me to construct an account of when and why, exactly, the ordinary stringency of an agent's fact-relative and evidence-relative duties against (risking) harm is reduced by the fact that he will be infringing these duties for the benefit of other people and/or the fact that he has an other-regarding justification for doing so. Any such account would need to answer several difficult questions. What if the agent is acting partly for the benefit of others, but also partly for his own benefit? What if his justification for acting is partly other-regarding, but partly self-regarding? What if the agent is acting to benefit people to whom he owes special moral duties (such as his children)? Is it important that the agent would be permitted to refrain from engaging in his altruistic course of action (in light of the cost to him, or others, that doing so involves)? Is it important that the agent is acting for the benefit of the very people whose rights against harm he is about to infringe? Is it important that the agent is *motivated* by the expected benefits he is providing to other people, or might it suffice that he is *aware* of these benefits?

The case-based work of this section has implicitly suggested some answers to these general questions. But these implicit answers provide nothing like a complete account of this philosophical terrain.

---

[139] Haque, *Law and Morality at War,* p. 76.

[140] Why not ground strict defensive liability directly in the stringency of fact-relative duties against harm, as Haque suggests? See above, pp. 26 – 30.

The questions are important, moreover, quite apart from any concern with strict defensive liability: they bear on first-personal deliberation about the moral permissibility of harmful action. My limited claim is that an altruistic threatener's strict defensive liability (or lack of it) will also depend on the correct answers to these questions.

Nevertheless, let me conclude this section by offering a final, tentative argument in favor of my general stance about these issues, as against Hosein and Haque's. As I have said, it seems especially implausible that a threatener should be liable to defensive killing when he is reasonably risking harm to the very victim whom his action promises to benefit *ex ante.* That is the case in Altruistic Injection. I would be more hesitant to abandon this sort of judgment than to abandon the judgment that altruistic threateners do not incur strict defensive liability in cases like Fighter Pilot (cases in which the relevant threatener is conferring an *ex ante* benefit on different people than the people on whom he is imposing risks of harm). But the distinction between these two kinds of cases, while clear enough on first inspection, may become fraught in other scenarios. Consider

> *Big Opaque Box:* A big opaque box contains twenty men, nineteen of whom are dying. Threatener is about to release a gas into the box. As she knows, releasing the gas will save the nineteen men who are dying and kill the one remaining man (who will otherwise live). Threatener knows that the nineteen men who will be saved are wearing red shirts and the one man who will be killed is wearing a blue shirt. She also knows the given names of all twenty men ("Abraham, Benjamin, Carlos…").

Whether Threatener's action is in the *ex ante* benefit of the man whom it is about to kill depends on how that person is described. On certain descriptions ("the man who is wearing a blue jersey," "the man who will in fact be killed if the gas is released"), releasing the gas into the chamber is not in this man's benefit *ex ante.* By contrast, if this man is picked out by his given name, then releasing the gas *is* in his benefit *ex ante.* For it is true that, of any given man so described ("Abraham" or "Benjamin" or "Carlos" or…), releasing the gas is 95% likely to save him from being killed and 5% likely to kill him.

If we distinguish morally between cases (such as Altruistic Injection) in which the relevant threatener's action is in the *ex ante* benefit of the very person who she is threatening to harm and cases (such

as Fighter Pilot) in which the relevant threatener's action is not, then our verdict about the threatener's defensive liability in this case will require us to settle upon a relevant description. Not only does that seem a bit strange, but formidable difficulties may arise in attempting to provide an adequate general account of which descriptions are relevant. That is the lesson of some fascinating recent work on the morality of risk imposition.[141] Adopting a uniform treatment of cases like Altruistic Injection and cases like Fighter Pilot would allow our account of strict defensive liability for altruistic wrong to circumvent any such difficulties. And since it is extremely plausible that strict defensive liability does not attach in the former sort of case, the best considered stance may be that it does not attach in *either* sort of case.

This sort of argument may have only limited force. As I have said, the relevant difficulty (or something closely akin to it) arises in a nearby philosophical domain, and so we are independently required to find some solution to it. Still, it is worth considering how to respond if - as seems quite possible - the general difficulty cannot be satisfyingly addressed. In the present context, I believe, the best response is to embrace a general exemption from strict defensive liability for faultless threateners across cases like Altruistic Injection and Fighter Pilot.

VII. AVOIDABILITY AND LIABILITY

If one of two parties must suffer harm, it is often plausible to suppose that the party who had a better opportunity to avoid suffering harm should (other things equal) be the one to suffer it. Several philosophers have argued, accordingly, that facts about avoidability have an important role to play in grounding liability to defensive harm.[142] The basic idea is that a threatener is liable to defensive harm on behalf of his victim when and because the former party had a better opportunity to avoid suffering harm than the latter. This *avoidability account* may be related to the idea, which I have criticized, that defensive liability is grounded in

[141] See especially A. Mahtani, "The *Ex Ante* Pareto Principle," *The Journal of Philosophy* 114 (2017) 6: 303 - 323.
[142] N. Kolodny, "Political Rule and its Discontents," in *Oxford Studies in Political Philosophy*, vol. 2, eds. D. Sobel, P. Vallentyne, and S. Wall (Oxford: Oxford University Press, 2016): 35 - 68; T.M. Scanlon, "Thomson on Self-Defense," in *Fact and Value,* eds. A. Byrne, R.C. Stalnaker, and R. Wedgwood (Cambridge: MIT Press, 2001); Tadros, *The Ends of Harm,* esp. pp. 169 - 197.

luck egalitarian distributive fairness.[143] But it can also be understood and defended in other ways, and some of the view's most prominent proponents, such as Thomas Scanlon, are keen to do so.[144] So it is worth considering the merits of the avoidability view at least partly in abstraction from these competing rationales.

In this section I will argue that considerations of avoidability have much less bearing on defensive liability - and strict defensive liability in particular - than such views envisage. Avoidability may have some limited role to play in undercutting strict defensive liability that a threatener would otherwise incur. But articulating the nature of this role will prove to be a fiendishly difficult task.

The strongest form of the avoidability view holds that a threatener will become liable to defensive harm only when, and solely because, he enjoyed a better opportunity to avoid suffering harm than his victim.[145] On this view, all forms of defensive liability - strict defensive liability included - are grounded entirely in considerations of avoidability. But this strong view is untenable. Other considerations, like a threatener's culpability, have an important and irreducible role to play in grounding defensive liability. To illustrate the point, suppose that a mobster credibly threatens to break Isabel's legs unless she breaks the legs of the local district attorney (who is currently prosecuting the mobster). Isabel is unfazed by this threat, however, for she was *already* planning to break the legs of the district attorney out of sheer malice: the district attorney happens to be her rival in a tempestuous love triangle. At the same time, the mobster credibly threatens to break the legs of a different person, Clarice, unless *she* breaks the legs of the district attorney.

---

[143] As typically understood, luck egalitarian distributive fairness turns partly on whether the harm which an agent suffers can be suitably attributed to his responsible choices. And one natural thought is that an agent's responsible choices will have such a legitimating role when and because (a) he had adequate opportunities to avoid the harm in question and (b) he responsibly failed to avail himself of these opportunities.

[144] According to Scanlon, avoidability bears on liability because we are less likely to suffer harm under moral principles that make our protections against being harmed sensitive to our opportunity to avoid harm. T.M. Scanlon, *What We Owe to Each Other* (Cambridge: Belknap Press, 1998): 248 - 294. According to Niko Kolodny, avoidability bears on liability because of our interest in having *control* over whether we suffer harm. Kolodny, "Political Rule." According to Stephen Perry, avoidability bears on liability because it bears on agent responsibility for outcomes. S.R. Perry, "Responsibility for Outcomes, Risk, and the Law of Torts," in *Philosophy and the Law of Torts,* ed. G.J. Postema (Cambridge: Cambridge University Press, 2001): 72 - 130. Scanlon's rationale (and Kolodny's, too, perhaps, depending on how it is interpreted) seems incapable of underwriting any of the intuitive distinctions that (I argue below) an adequate account of avoidability's bearing on liability must explain. Perry's may be more promising.

[145] This position is explicitly embraced or at least strongly intimated in Kolodny, "Political Rule"; Scanlon, "Thomson on Self-Defense"; and Tadros, *The Ends of Harm.* Tadros has recently abandoned it, partly on the basis of cases like the one I offer in the text above. See *To Do, To Die, To Reason Why,* pp. 146 - 155.

Clarice accedes to this threat solely in order to preserve her legs. Both Isabel and Clarice are liable to significant defensive harm, if necessary, to prevent them from harming the district attorney. But it seems clear that an impartial observer should harm Isabel rather than Clarice, if he can protect the district attorney equally well by harming either assailant (and cannot divide the necessary harm between them). Moreover, it seems plausible that Isabel will be liable to a greater maximum amount of harm than Clarice will be.[146]

These judgments would be inexplicable if defensive liability were grounded solely in facts about avoidability, for Isabel enjoys no better an opportunity to avoid harm than Clarice. Instead, it seems clear, these two assailants are morally distinguished by their differing levels of culpability for the threats they pose. Insisting on avoidability as the sole determinant of defensive liability would force us to abandon such judgments. To be sure, if culpability lacked other kinds of moral significance, there might be some pressure to revise our judgments in this way. But culpability is pervasively significant in ordinary moral thought, quite apart from questions about liability to harm; it bears, for example, on the fittingness of blame, resentment, and forgiveness. So there is no good reason to subject our intuitive convictions about defensive liability to any such procrustean butchery, by explaining culpability's significance entirely in terms of avoidability and abandoning judgments that do not fit within this narrow mold.

A more modest and plausible view is that avoidability explains the existence and scope of *strict* defensive liability: when an agent *faultlessly* threatens to inflict harm on another person, the threatener will be liable to defensive harm on behalf of his victim latter just in case he enjoyed a better opportunity to avoid harm than she did. But even this more modest view is too strong. Consider

> *Building Blaster:* Unless he makes a lot of money, Threatener will almost certainly suffer financial ruin within a year, in virtue of enormous debts that will soon come due (and which he had no good opportunity to avoid incurring). Luckily, he discovers an abandoned and unowned plot of land right next to his town's main public park. The land is currently occupied by a dilapidated and monstrous

[146] For example, it might well be permissible to paralyze or even kill Isabel, in order to prevent her from breaking the district attorney's legs, if absolutely necessary to do so. It would seem impermissible to paralyze or kill Clarice in order to prevent *her* from breaking the district attorney's legs, even if this were absolutely necessary to do so.

old building. If Threatener destroys this building - which will require several months of blasting it with high-powered explosives - he can then build a successful commercial establishment on top of the land and avoid financial ruin. As he knows, however, there is some small chance that, at some point over the course of the next few months, this blasting work will send sharp debris into the nearby park, killing a person in it. But this risk is sufficiently small that Threatener is permitted to create it. Threatener posts signs up all over town letting the townspeople know that, if they use the town's main pubic park, there is a very small chance that they might be killed by debris from his blasting. (This risk is sufficiently small that it is rational for townspeople to continue using the park.) But one day this risk materializes: Threatener's blasting is about to kill Victim, who is walking her dog in the park, and an impartial observer can save Victim only by defensively killing Threatener.

It seems plausible that Threatener is liable to be defensively killed, in order to prevent him from killing Victim through his blasting. But Threatener did not have a good opportunity to avoid creating this threat: the cost of refraining from creating it was a near certainty of financial ruin. Victim had a better opportunity to avoid becoming subject to this threat of harm: she could have done so by refraining from walking her dog in the park for a couple of months. So the cost to Threatener, of avoiding posing this threat, was greater than the cost to Victim of avoiding it. If comparative avoidability is the basis of strict defensive liability, therefore, Threatener is *not* liable to defensive killing in this case.

That seems implausible. By blasting in order to build his hotel, Threatener is creating serious risks of harm to people in the park. It seems that he would be obligated to internalize these risks if he could: if Threatener could blast in a way that imposes small risks of death on himself, instead of blasting in a way that imposes small risks of death on people in the park, he would be obligated to blast in the former way. Moreover, it does not seem that Threatener can relieve himself of these risk internalization obligations simply by giving people who wish to use the public park prior notice of his blasting. If that were so, Threatener would in effect have a power to declare that anyone who uses this park over the next few months does so at her own peril. The better view is that Threatener would be obligated to internalize the risks of death that he is imposing on park users, if he could, and that he is liable to defensive measures that will in

effect force him to internalize this risk. And that is it true despite the fact that "the costs of avoiding the threat-creating activity are much greater than the costs of avoiding being exposed to the threat."[147]

The cases we have already considered, in the preceding sections of this article, supply further evidence against the comparative avoidability account. Consider again the threateners in Sickly Driver and Mistaken Attacker. Each threatener can avoid suffering a near certainty of death only by risking killing another innocent person. Neither threatener has any good opportunity, therefore, to avoid creating this risk of harm, or becoming liable to defensive harm in consequence; each threatener's opportunity to avoid harm is just as bad as his victim's. Nevertheless, it seems intuitively, each threatener is liable to defensive harm. And this makes sense, of course, on the picture of strict defensive liability that I have been developing. Through no choice of his own, an agent can find himself saddled with a highly stringent duty against risking harm to another person. If the agent suffers some harm in order to refrain from infringing this duty, the agent will have lacked a good opportunity to avoid this harm. So complying with our primary duties against risking harm can consign us to suffer harm which we had no good opportunity to avoid. And I have argued that strict defensive liability is a sort of functional surrogate for compliance with such duties. It makes sense, therefore, that an agent can become strictly liable to defensive harm even when he had no good opportunity to avoid suffering harm - no better an opportunity than his victim.

Thus the comparative avoidability account of strict defensive liability ought to be rejected: just as the stringency of a threatener's duties against (risking) harm to another person is not a simple function of the parties' respective opportunities to avoid harm, neither is agent's strict defensive liability. Even so, it remains possible that the stringency of an agent's duties against risk imposition - and derivative moral phenomena, such as his obligations to internalize risk and his strict defensive liability - are *sensitive* to his victim's opportunity to avoid suffering harm. Put differently, facts about avoidability might *undercut* strict defensive liability that a threatener would otherwise incur. By way of illustration, consider

[147] Tadros, *The Ends of Harm*, p. 231.

> *Train Driver:* Threatener is a train driver. He will almost certainly suffer severe harm unless he drives his train along a certain track, continuously and without interruption, for the next hour. Unfortunately there is another (adult and cognitively competent) person, Victim, currently playing on the track. If Threatener keeps driving his train, and Victim does not get off the track, the train will almost certainly crush and kill her. Victim can get off the track easily and without incurring any cost. However, as Threatener knows, Victim almost certainly will *not* get off the track in time to save her life, because she is reckless and wishes to keep playing on the track as long as possible.

By driving down the track, Threatener is creating a risk that he will hit and kill Victim - an extremely high risk, since she is (as he knows) extremely reckless and unlikely to get off the track in time. Yet it still seems permissible for Threatener to drive down the track. Even if he can internalize this risk, moreover, it does not seem that Threatener is obligated to do so. Suppose that he can install a device on his train that will blow it up and kill him if and only if it is going to hit and kill Victim. It does not seem to me that Threatener is obligated to install this device. Nor would he be obligated to request that a suitably situated impartial observer blow up his train, killing him, if necessary to save Victim. Nor may an impartial observer, on her own initiative, kill Threatener to save Victim: he is not liable to be defensively killed.

So Threatener's duty against risking death to Victim - indeed, imposing on her a near certainty of death - appears to be much less stringent, in this case, than such duties typically are. And the most natural way to explain this fact is by reference to Victim's opportunity to avoid suffering this harm. If Victim acts with due care for her own interests, by simply getting off the track, she can easily avoid being killed by the train. It is because she can easily avoid suffering harm, in this way, that Threatener's duties against risking such harm to her do not possess their ordinary stringency. And this is true, it seems to me, although it is an entirely foreseeable consequence of Victim's reckless nature that she will decline to avail herself of this opportunity and be killed as a result. So an agent's duties against risking harm to another person can be substantially weakened, it appears, by his victim's opportunity to avoid suffering such harm.

This limited proposal about the significance of avoidability is by far the most plausible one we have considered. But it must be interpreted with care, on pain of generating implausible results about other cases. Consider

> *Bomb Planter:* Victim is a passionate sportswoman, who has long wanted to ski down a certain beautiful mountain. Until now, however, she has refrained from doing so, for the mountain is famously perilous: by skiing down it, she will almost certainly trigger a massive avalanche and get crushed to death by it. One day, however, Victim throws caution to the wind and decides to ski down the mountain anyway. Miraculously, she is going to survive this imprudent action, and arrive at the bottom of the mountain unharmed. But now a villain, who wishes to see Victim dead, credibly threatens to inflict severe harm on Threatener unless he plants a device half-way down the mountain, which will almost certainly blow up and kill Victim when she passes by it. Threatener's choice, therefore, is between almost certainly suffering death himself and almost certainly inflicting death on Victim. (Threatener is aware that, if he does not plant this device, Victim will survive her trip down the mountain, against all the odds.)

Threatener's duty against imposing a near certainty of death on Victim, in this case, *does* seem highly stringent. Thus it seems impermissible for Threatener to infringe this duty, even in order to save himself from a near certainty of death.[148] And the stringency of this duty does not seem appreciably diminished by

---

[148] The case does not directly address strict defensive liability, of course. For if Threatener *does* choose to install the device, and thereby threatens to kill Victim, she will be culpably responsible for this threat and liable to defensive harm on that basis. Consider, however, consider the following variant of the case:

> *Bomb Planter (Small Risk Variant):* In order to get a cheap thrill, Victim decides to ski down an extremely steep mountain, knowing that doing so will very likely kill him. Miraculously, however, he is going to survive. But now a villain, who wishes to see Victim dead, credibly threatens to kill Threatener unless she plants a bomb half-way down the mountain. When Victim skis past this bomb, there is a 1% chance it will explode and kill him.

Here it is permissible for Threatener to install the device, thereby imposing a 1% risk of death on Victim: she may (evidence-relative) permissibly infringe her duty against imposing this risk on Victim. But this evidence-relative duty against risk imposition nevertheless seems to be a stringent one. Thus Threatener would be obligated to internalize this risk, if she could. For example, if Threatener could save herself from the villain by planting the device near herself - thereby risking killing herself rather than risking killing Victim - she would be obligated to do so. Similarly she would be obligated to request that an impartial observer kick the device back at her if this should prove necessary to save Victim. Similarly, I believe, she is liable to this defensive measure even if she cannot request it (because she cannot communicate with observer or is unaware of his presence).

the fact that Victim has an excellent opportunity to avoid suffering this harm. After all, Victim can easily preserve her life simply by refraining from skiing down the mountain, something she has no strong interest in doing. By contrast, Threatener can avoid inflicting a certainty of death on Victim only by allowing himself to suffer a near certainty of death.

The intuitive asymmetry between cases like Train Driver and cases like Bomb Planter needs to be explained. It is tempting to do so in the following way. In the former case, the victim has an excellent opportunity to avoid *the particular risk of harm* to which Threatener is subjecting her. In the latter case, the victim has an excellent opportunity to avoid a *different* risk of harm. Perhaps, therefore, the stringency of a threatener's duty against imposing some risk of harm on a victim can be reduced if (but only if) the victim enjoyed a sufficiently good opportunity to avoid becoming subject *to that very risk of harm.* The threatener cannot impose on the victim a risk of being harmed in some different manner from the manner she has reason to expect. As Holmes remarked about a related matter in law, with characteristically oblique rhetorical panache, "choice which entails a concealed consequence is as to that consequence no choice."[149]

This Holmesian proposal seems promising in the abstract. But it too encounters difficulty in concrete cases. Consider

> *Bomb Planter*:* As in Bomb Planter. However, the bomb that Threatener places halfway down the mountain will kill Victim by triggering an avalanche that will crush her (rather than by blowing her up, as in the original case).

Intuitively, this variant of Bomb Planter seems to me relevantly analogous to the original case. In both cases, it appears, Threatener has a stringent duty against imposing a near certainty of death on Victim by planting the bomb halfway down the mountain, and the stringency of this duty does not seem appreciably reduced by the fact that Victim can easily preserve her life simply by refraining from skiing down the mountain. But in Bomb Planter*, Victim will be killed by an avalanche, *just as she foresees*. Threatener is thus imposing on Victim a risk of being killed in the very same way that, if she skiis down the mountain, she foresees she will

[149] Cf. O.W. Holmes, *The Common Law* (Cambridge: Belknap Press, 2009), 86.

incur. So it would seem that Victim has an excellent opportunity to avoid becoming subject to the very risk of harm – i.e. of being killed by an avalanche - to which Threatener is subjecting her.

The Holmesian proposal is evidently deficient, therefore, in that it fails to secure the intuitive distinction between cases like Train Driver (on the one hand) and Bomb Planter and Bomb Planter* (on the other). There is another reason to resist the Holmesian proposal. Compare these two cases:

> *Release the Lion:* Threatener will almost certainly be murdered by a whimsical villain unless he releases a deadly lion into a certain deserted grove. Threatener warns Victim that she should not enter the grove, because it contains a deadly lion who will almost certainly maul her to death if she goes near it; then he releases the lion into the grove. Victim enters the grove in order to see the lion and she is promptly mauled to death by it.
>
> *Release the Deadly Bear:* Threatener will almost certainly be murdered by a whimsical villain unless he releases a deadly bear into a certain deserted grove. Threatener warns Victim that she should not enter the grove, because it will soon contain a deadly *lion* who will almost certainly maul her to death if she goes near it. (The whimsical villain allows Threatener to falsely warn Victim that there will be a lion in the grove, but he forbids Threatener to accurately warn Victim that a bear will soon be in the grove. The villain is very whimsical indeed.) After making this warning to Victim, Threatener releases the bear into the grove. Victim enters the grove in order to see the lion and she is promptly mauled to death by the bear.

On the Holmesian proposal, Threatener's duty against imposing a near certainty of death-by-lion-mauling on Victim has greatly diminished stringency in Release the Deadly Lion. For Victim has an excellent opportunity to avoid becoming subject to this very risk of death: she has reason to believe that she will almost certainly be killed by a bear if she enters the grove. By contrast, Victim does *not* have an excellent opportunity to avoid becoming subject to the very risk of death in Release the Deadly Bear: she has reason to believe that she will almost certainly be killed by a *lion* if she enters the grove, whereas in reality Threatener is imposing on her a risk of being killed by a *bear.* But it seems strange, intuitively, to distinguish between the cases in this way.

We might attempt to address the problem by characterizing the risk to which Threatener subjects Victim, in both cases, at a higher level of generality: it is a risk of *being mauled by a deadly animal,* or *being killed by a deadly animal,* or some description along those lines.[150] But once we are in the business of generalizing away from the details of each particular threat, in this way, there seems to be no principled reason to stop short of generalizing away *all* the particular details - except for the magnitude of the relevant risk, and the severity of the harm that is risked. It seems opaque, after all, why any description pitched at an intermediate level of generality should be morally relevant; neither threatener nor victim has any reason to care whether the relevant threat takes the form of a lion, or a bear, or a tornado, or a bomb. But to accept that we should describe risks with such austere generality would eviscerate the motivation for entertaining the Holmesian proposal in the first place. If (for example) we characterize the risk to which Threatener is subjecting Victim in Bomb Planter simply as "a near certainty of death," then Victim does indeed enjoy an excellent opportunity to avoid this very risk: she knows that she will almost certainly die if she skiis down the mountain. In this and all other cases the Holmesian proposal would lose all content.

There are, of course, other conceivable strategies by which we might seek to distinguish cases like Train Driver from cases like Bomb Planter. For example, in the former case the threatener does not just create a risk of death to his victim; his activities also *make her aware* that she will incur a risk of death unless she takes suitable action. In the latter case, by contrast, the victim acts in awareness that she is about to incur a risk of death, but this awareness arises quite independently of the threatener's activities. Perhaps a threatener's duty against imposing a risk of harm on his victim is diminished in stringency only if the threatener's own activities are responsible for her awareness of this risk: the threatener must *place his victim in a good position* to extricate herself from the threat to which he is subjecting her.

Like the other strategies we have considered, this idea has some abstract appeal, but it encounters formidable resistance from concrete cases. Consider

[150] A similar maneuver is often employed to address a similar problem that confronts proximate cause or "scope of the risk" analysis in tort law. See H.M. Hurd and M. Moore, "Negligence in the Air," *Theoretical Inquiries in Law* 3 (2002) 2: 333 - 411. This maneuver is open to an objection that closely parallels the objection that I am pressing in the text above.

Release the Lion (Messenger Variant): In order to avoid being murdered by a whimsical villain, Threatener releases a deadly lion into a certain deserted grove. Then he sends his messenger to inform Victim that she should not enter the grove, for she will almost certainly be mauled by the lion if she does. As Threatener knows, there is a 50% chance that his messenger's car has broken down, so that the warning has not been delivered. But then Threatener learns that - as it just so happens, in a feat of incredible coincidence - Victim's rascally friend Frederic has convinced her that there is a deadly lion in the deserted grove, which will almost certainly maul her if she goes in. Frederic has done this as a practical joke; neither the presence of the lion, nor anything else causally traceable to Threatener's activities, has led him to play this trick on Victim.[151]

It seems implausible that the stringency of Threatener's duty against risking death to Victim should depend on whether his messenger has been intercepted. It would seem unduly self-involved for Threatener to care much about this fact, rather than whether Victim *has come to believe* in the existence of the risk of which the messenger is seeking to inform her. The causal genesis of this belief does not seem pertinent to whether Threatener may permissibly impose the relevant risk on Victim, or whether he must internalize it, or whether he is liable to defensive harm that will in effect force him to do so.

So I am at a loss to articulate how, exactly, the stringency of a threatener's duty against risking harm to another person depends on the threatener's opportunity to avoid suffering harm. It is powerfully intuitive that, in cases like Train Driver, avoidability has some significance of this general kind. But it is equally plausible that the nature of the victim's opportunity to avoid harm in cases like Bomb Planter does *not* undercut the stringency of the relevant threatener's duty against imposing risk on her. An adequate set of moral distinctions between these and other pertinent cases is elusive. Still, it is not unreasonable to hope that with sufficient reflection an adequate set of distinctions will be found. Philosophical discussions of avoidability's bearing on liability have tended to proceed at a high level of generality - highlighting cases that demonstrate that avoidability has *some* significant role in shaping liability, without attending to anything

[151] The structure of this case is closely analogous to Edmund Gettier's famous thought experiment about knowledge, presented in "Is Justified True Belief Knowledge?" *Analysis* 23 (1963) 6: 121 - 123.

like the full range of cases in which avoidability may or may not have such significance. But this broader inquiry is important. Scenarios at the periphery of everyday moral perception may reveal instabilities at its core. So it is with avoidability and liability: the existence of a significant connection between these phenomena is as intuitively compelling as it is difficult to articulate with precision.

The point is continuous with some more general observations about this article's aspirations and limitations. The article has sought to establish the foundations of an elegant explanatory structure. Whether a faultless threatener is liable to defensive harm, in order to prevent a risk she is imposing on her innocent victim from materializing, depends on the stringency of her evidence-relative duty against imposing the risk in question. That in turn depends, presumptively, on the stringency of her fact-relative duty against inflicting this harm. But the presumption can be undercut, by factors such the victim's consent to this risk imposition, the fact that the victim has reciprocally imposed an equivalent risk on the threatener, or the fact that the threatener is acting (entirely) for the benefit of other people and with (entirely) other-regarding justification, or the fact that this harm is suitably avoidable by the victim. Such factors may reduce the ordinary stringency of the threatener's duty against imposing the relevant risk on his victim - such that he is not obligated to internalize the risk, and is not liable to defensive measures that will in effect force him to internalize it.

This structure is inherently plausible and it has considerable explanatory power, but we have also seen how it inherits some of ordinary moral thought's more mysterious commitments. These include the insensitivity of our deontological moral duties to powerful concerns about luck egalitarian distributive fairness; the sensitivity of such duties to elusive distinctions between different kinds of opportunities to avoid harm; and their sensitivity to murky moral conceptions about property and normative powers. The general point does not make trouble merely for the particular explanatory account I have proposed. If anything like the extensional connections I have highlighted are sound, then all of the uncertainties, perplexities and disagreements that characterize our thinking about the nature and grounds of primary deontological duties can be expected to reproduce themselves in our thinking about strict liability to defensive harm.

It is unsurprising, therefore, that in places our investigation has amounted to little more than "[a]n expression of difficulties, uncertainties and doubts, arriving at no very definite conclusion".[152] The enterprise is sufficiently demanding, and its rewards sufficiently unclear, that we may be tempted to cast our lot with some different deontological denomination - one that disdains such philosophical labors as Jesuitical casuistry, and seeks instead to deliver its moral prescriptions about risk and defense from loftier theoretical perch.[153] But that would be a mistake. Ordinary moral thought cannot be divested of its more intricate and reflectively mysterious elements without doing considerable violence to its central commitments; it is the basic architecture of those commitments that is itself frequently intricate and mysterious. A philosophical procedure that overlooked these elements, or ruthlessly pared them down in the name of elegance or abstract appeal, would produce a deontological view that is unrecognizable to ordinary moral conviction.[154] And it is fidelity to ordinary moral conviction that justifies deontological moral theory, in my view, if anything does. Spiritual transformation may still be called for, of course. But if so I suspect it will be more radical in character: in my experience, at least, conversion to consequentialism is a standing hazard of searching deontology.

[152] W. Prosser, "*Palsgraf* Revisited," *Michigan Law Review 5*2 (1953) 1: 1 - 33, 2.
[153] The Kantian Church, long renowned for its learned discourses on corrective liability, has recently turned its attention to the morality of defense. See A. Ripstein, "Reclaiming Proportionality," *Journal of Applied Philosophy* 34 (2017) 1: 1 - 18.
[154] See, e.g., *id.* at p. 16, fn. 8., in which it is decreed that private persons are categorically forbidden from non-consensually imposing substantial risks of serious harm on other (non-liable) private persons.

STRICT LIABILITY TO DEFENSIVE HARM II: THE SPECTER OF ABSOLUTE LIABILITY

There is a close connection between obligations to internalize risks of harm and liability to defensive harm.[155] Suppose that a threatener imposes a risk of harm on another person, and the threatener would be obligated to internalize this risk if he could. Then the threatener is liable to a defensive measure that will, in effect, force him to internalize this risk. He is liable, moreover, even if he is acting entirely without moral, cognitive, or epistemic fault. His liability is, in this sense, "strict." Thus even a wholly impartial observer may defensively harm such a faultless threatener in order to prevent him from inflicting equivalent harm upon his victim. By contrast, if this faultless threatener's evidence-relative duty against risking harm to his victim is *not* so stringent that he would be obligated to internalize this risk, then he is not liable to defensive measures that will in effect force him to internalize it.

Obligations to internalize risk and strict defensive liability are closely connected, in this way, because both are grounded in the stringency of the evidence-relative *pro tanto* duty against risk imposition which the threatener is infringing. In turn, the stringency of this evidence-relative duty is presumptively grounded in the stringency of the fact-relative duty against harm whose infringement the threatener is risking. The presumption is not absolute, however: facts such as the victim's consent to the imposition of the risk in question, or (perhaps) the fact that she has reciprocally imposed such a risk on him, may undercut the ordinary stringency of his evidence-relative duty against imposing the relevant risk upon her.

In the predecessor to this article, I developed these abstract claims in conjunction with several case-based arguments, concerning both novel hypothetical examples (such as the Sickly Driver cases[156]) and familiar ones (such as Conscientious Driver[157] and Mistaken Attacker[158]). In this article I will deploy this framework in order to examine the defensive liability of agents who impose only astronomically small risks of harm on their victims. Such threateners are typically characterized as "non-responsible threats." This label reflects a widespread philosophical assumption that, if it is permissible to defensively kill such

[155] "Strict Liability to Defensive Harm I: Internalizing Risk and Justifying Defense."
[156] *Id.* at 11 – 30.
[157] *Id.* at 52 – 73.
[158] *Id.* at 38 – 41.

threateners, that is simply because of the threats their bodies pose. But reflection on the connection between obligations to internalize risk and strict defensive liability casts powerful doubt on this assumption. It provides reason to believe that such threateners render themselves liable to severe defensive harm through responsible exercises of their agency, not merely through the movement of their bodies.

We may be able to resist this counterintuitive conclusion, but the task will prove more difficult than we might expect. One popular strategy, which appeals to the idea that astronomically small risks of harm are *unforeseeable,* will prove entirely unfit for purpose. Other strategies may have greater promise. The first strategy appeals to the moral significance of reciprocal risk imposition. The second strategy appeals to the fact that an agent who imposes an astronomically small risk of harm on another person will not, in many circumstances, increase the overall amount of risk to which she is subject, relative to the baseline amount of risk to which she is subject simply in virtue of the agent's passive existence in the world. The third strategy invokes the fact that, when someone is exposed to an astronomically small risk of harm, he is typically provided with an astronomically small expected benefit as well.

Each strategy warrants close consideration and further development. For each strategy identifies a kind of moral consideration that may indeed bear on the stringency of duties against risk imposition (and so bear on obligations to internalize risk and strict defensive liability). But none of these strategies, I will argue, can establish a categorical exemption from strict defensive liability for threateners who impose only astronomically small risks of harm. The more we reflect on the relevant range of cases, moreover, the less plausible any such categorical exemption becomes. And I will offer, in conclusion, three novel and mutually reinforcing arguments for thinking that astronomically small risk imposition can indeed incur strict defensive liability. There may be no safe harbor from liability to severe defensive harm: the possibility of incurring defensive liability may be an inescapable risk of rational agency.

I. The Cast of Characters

The kinds of threateners that concern me will be familiar to anyone who has seriously engaged with the contemporary philosophical literature on defensive harm. Here are the standard examples:

Phone Operator: Threatener is about to call his favorite restaurant on his cell phone, in order to place an order for delivery. Unbeknownst to him, a villain has programmed his phone such that dialing it will detonate a bomb that will destroy a neighboring house, killing the person (Victim) inside it. Observer knows this, and can save Victim only by defensively shooting and killing Threatener with a sniper rifle, before he dials his cell phone.[159]

Light Switcher: Threatener reaches for the light switch on his bedroom wall. Unbeknownst to him, flipping the switch will blow up Victim's house and kill her (because of an astronomically improbable electrical current between his house and hers). Observer knows this, and can save Victim only by shooting and killing Threatener before he flips the switch.[160]

Falling Man: Threatener is taking a stroll in a grove, when a freakishly powerful gust of wind picks him up and throw him down a well. His body is now about to crush and kill another innocent person, Victim, whom the wind previously deposited there. Observer knows this, and can save Victim only by defensively killing Threatener. (Threatener himself will survive the fall, thanks to the cushion provided by his ample belly.)[161]

There is disagreement, in each case, about whether the victim would be permitted to kill the agent who threatens her, if necessary to save her own life. Most philosophers say yes; a few say no.[162] There is much more disagreement about whether impartial observers may defensively kill such threateners. Among philosophers on both sides of this debate, however, it is widely assumed that - *if* it is permissible to kill these threateners - that is simply because of the threats their bodies pose.[163]

---

[159] Closely modeled on. J. McMahan, "The Basis of Moral Liability to Defensive Killing," *Philosophical Issues* 15 (2005): 386 - 405, 397.
[160] Closely modeled on J. Thomson, *The Realm of Rights* (Cambridge: Harvard University Press, 1990): 229.
[161] Closely modeled on R. Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1997): 34
[162] See e.g. M. Otsuka, "Killing the Innocent in Self-Defense," *Philosophy & Public Affairs* 23 (1994) 1: 74 - 94.
[163] There are notable exceptions. See e.g. L. Alexander, "Recipe for a Theory of Self-Defense: The Ingredients, and Some Cooking Suggestions," in *The Ethics of Self-Defense,* eds. C. Coons and M. Weber (Oxford: Oxford University Press, 2014): 20 - 50, 28; K. Ferzan, "Culpable Aggression: The Basis for Moral Liability to Defensive Killing," *Ohio State Journal of Criminal Law* 9 (2013): 669 - 95, 682; A. Walen, *The Mechanics of Claims and Permissible Killing in War* (Oxford: Oxford University Press, 2020): 157.

Indeed, as I have observed, this assumption appears to be reflected in the standard philosophical label, "non-responsible threat", that is used to refer to such characters. These "non-responsible threats" are typically contrasted with the threateners in cases like Mistaken Attacker and Conscientious Driver, who are commonly characterized as "minimally responsible threateners."[164] These labels are philosophical terms of art, and of recent provenance. So there is no canonical definition of either label. But the basic idea is that minimally responsible threateners pose *reasonably foreseeable* threats of harm: while the risks they are imposing on other people are fairly small, these risks are sufficiently large (or of a suitable character) that the resulting harms can fairly be foreseen. The harms imposed by "non-responsible threats," by contrast, are *not* reasonably foreseeable. If it is permissible to defensively kill such threats, therefore, the explanation has nothing to do with the moral significance of their responsible actions. It is only because their bodies are moving in ways that threaten to harm other people.

I will now argue that this assumption is misguided. There is considerable philosophical pressure to conclude that the action of each "non-responsible threat," rather than (or in addition to) the mere movement of his body, renders him liable to be defensively killed. Each threatener is about to infringe a stringent fact-relative duty against killing another person. Presumptively, therefore, each threatener has a stringent evidence-relative duty against imposing any given risk of death on the other person. Unless the presumptive stringency of this duty against risk imposition is undercut, the threatener is obligated to internalize this risk (if he can). And even if he cannot internalize the risk, he is liable to a defensive measure that will in effect force him to internalize it. This line of thought will be familiar to those who have read the predecessor of this article. There, I argued that this sort of *risk internalization argument* delivers plausible verdicts about strict defensive liability in many cases and illuminates their moral structure. My claim here is that this form of reasoning is just as compelling when the relevant risks are astronomically small.

[164] See e.g. S. Bazargan, "Killing Minimally Responsible Threats," *Ethics* 125 (2014) 1: 114 - 136; McMahan, "The Basis of Moral Liability."

Let us consider one of the relevant cases, Phone Operator, in detail. The first order of business is to establish that the threatener in this case is about to infringe a highly stringent fact-relative duty against killing another person. Here is a variant of the case that makes the matter clear:

> *Phone Operator With Urgent Call:* Threatener is suffering from a deadly heart attack and urgently needs to call the local ambulance service in order to take him to the hospital. But unfortunately, as he knows, a villain has hacked into his phone and programmed it, so that dialing it will blow up the house of his neighbor, Victim, and kill her.

It seems impermissible, in both the fact-relative and evidence-relative senses, for Threatener to dial his phone. His fact-relative duty against killing Victim is sufficiently stringent that he is fact-relative obligated to let himself die, rather than infringe this duty.

Threatener's choice in the preceding case is between lethal inaction (letting himself die) and lethal action (killing Victim). But the stringency of the relevant fact-relative duty against killing will also be significant in choices between two lethal actions. Consider

> *Phone Operator With Urgent Call*:* Threatener can save himself from a fate far worse than death either by (i) dialing a phone number that will blow up his house, killing him, or (ii) dialing a phone number that will blow up the house of his neighbor, Victim, killing her.

If the fate far worse than death which he stands to suffer is sufficiently terrible, Threatener would be permitted to blow up and kill his neighbor as a necessary means of escaping this fate. But in the case at hand, he can escape this terrible fate by blowing up and killing himself instead. It seems clear that he is fact-relative obligated to do so. These further cases thus exhibit the considerable stringency of the fact-relative duty - against blowing up another innocent person's home and killing her - that Threatener is about to infringe in the original Phone Operator. This duty seems just as stringent as other fact-relative duties against side effect killing (such as the duty against hitting and killing another person with a car, or the duty against shooting and killing another person with a gun) typically are.

The stringency of Threatener's fact-relative duty against blowing up and killing Victim will presumptively be reflected in the stringency of his evidence-relative duty against imposing any given *risk* of

this harm upon her. Absent some undercutting moral consideration, Threatener will be obligated (when acting for his own benefit) to internalize this risk if he can. Consider

> *Phone Operator With Urgent Call and Risk-Internalizing Button:* Threatener is suffering from a deadly heart attack and urgently needs to call the local ambulance service in order to take him to the hospital. As he knows, however, a villain might have hacked into his phone and programmed it - such that dialing the phone will detonate a bomb, blow up his neighbor Victim's house, and kill her. By calling the local ambulance service, therefore, Threatener will impose a 1% risk of death on Victim. However, Threatener can press risk-internalizing button before he dials the phone. If Threatener does this, then by dialing the phone he will instead create a 1% risk of blowing up his own house and killing himself.

Threatener is obligated to internalize the 1% risk of death he is creating for his own benefit, by pressing the risk-internalizing button. This position is both intuitively plausible and entailed by the presumptive parity principle we formulated in the predecessor to this article.[165] But now consider a case where the relevant risk of death is much smaller:

> *Phone Operator With Urgent Call and Magical Risk-Internalizing Button:* Threatener is suffering from a deadly heart attack and urgently needs to call the local ambulance service in order to take him to the hospital. Threatener has read the latest philosophy journals, and he knows that all human action is pregnant with risk. So he knows there is some astronomically small risk that, by dialing his phone, he will infringe a stringent fact-relative duty against killing another innocent person. However, Threatener has a way to internalize all such risks: he can press a magical button, before he dials his phone. Then the phone will blow up, killing Threatener, if and only if he is about to infringe a stringent fact-relative duty against killing another person by dialing it.

Is Threatener obligated to press this button before dialing his phone? I am intuitively inclined to say no. My intuition is weak, however, and it may be doubted whether our intuitions about such fantastical cases are

[165] "Strict Liability to Defensive Harm I," p. 15.

reliable. But the case is revealing even if we discount the evidentiary force of such intuitions entirely. The critical point involves the *structure* of the case.[166] Threatener is risking the infringement of a stringent, fact-relative duty against killing another person. And the case is *precisely analogous* to any other case in which an agent can choose to internalize such a risk - apart, of course, from the astronomically small (or freakishly unusual) character of the risk involved. Absent some principled explanation of why that should matter, therefore, we are under reflective pressure to accept that Threatener *is* obligated to internalize the risk in this case, just as Threatener was obligated to internalize 1% risk of death in the previous case.

If Threatener is indeed obligated to internalize this astronomically small risk, if he can, then Threatener is also liable to a defensive measure that will in effect force him to internalize this risk, even if he cannot (because he has no magical button or comparable means of risk internalization available). For suppose that Threatener can internalize the risk, not by pressing a magical button, but by asking a suitably situated impartial observer to monitor him and defensively kill him (if necessary) in order to prevent him from killing people such as Victim. If Threatener was obligated to internalize the risk by pressing the magical button, he is surely obligated to internalize it by requesting this functionally equivalent defensive course of action. And if he would be obligated to request this course of defensive action, then he is liable to suffer it even if he *cannot* request it (because he is physically unable to communicate with the impartial observer, or unaware of her existence). This defensive measure will, in effect, force him to internalize a risk he would be obligated to internalize if he could.

The preceding argument closely tracks the risk internalization argument about the Sickly Driver cases that I developed in the predecessor to this article.[167] That argument is more fleshed out, of course, but it would add little value to flesh out this argument in equal (and precisely parallel) detail. My point is that this argumentative structure is just as plausible in the present context - where we are dealing with astronomically small risks of harm - as in more familiar situations. Each inference in the argument is

[166] It is relevant, moreover, that this magical button is simply the *ex ante* equivalent of a risk internalization device that is ubiquitous in modern life: the ordinary commercial contract. And such contracts can, like this magical device, force agents to bear risks that are astronomically small, or even genuinely unforeseeable. See below, pp. 108 – 109.
[167] "Strict Liability to Defensive Harm I," pp. 11 – 30.

compelling. And these inferences are underwritten by the inherently plausible extensional principles - the presumptive parity principle and the risk internalization principle - that this article's predecessor formulated and deployed.[168] Indeed, those principles are appealing partly because they capture the structural logic of precisely such inferences.

Of course, insofar as such inferences threaten to saddle us with an unwelcome conclusion about strict defensive liability for astronomically small risk imposition, we might wish to revisit them. In particular, we might reject any close extensional connection between obligations to internalize risk and strict defensive liability. On this view, each agent *would* be obligated to internalize the astronomically small risks of harm he is imposing on other people, if he could, but he is not liable to defensive measures that will in effect force him to internalize such a risk.

It is worth constructing another variant of Phone Operator to bring out just how bizarre any such position would be. Suppose that Threatener is standing face to face with an impartial Observer. As both parties know, Observer will somehow learn - after Threatener starts dialing his phone, but before he finishes doing so - whether Threatener will thereby detonate a bomb and kill another person. On the position under consideration, the parties are obligated to attempt to communicate with one another, in order that Threatener may request that Observer defensively monitor and kill him (if necessary) to prevent him from killing another person in this way - but if the parties *cannot* communicate with each other, because (say) their mouths happen to be taped shut, then Observer is morally precluded from engaging in this very course of action. This conjunction of views strikes me as bizarre, as the conceit of locating Threatener and Observer face-to-face vividly exhibits. But the basic point - that there is an intimate connection between Threatener's obligations to internalize the risks he is creating and Observer's permission to make him internalize these risks - holds quite independently of where these characters are located, or what knowledge the former might have of the latter.

[168] *Id.*

In my view, therefore, we should maintain that - in all cases, including cases involving astronomically small risks of harm - a threatening agent is liable to harmful defensive measures, that will in effect force him to internalize a risk of harm, just in case he would be obligated to internalize that risk if he could. And this is just as true, I believe, when the threatening agent is acting for the sake of some trivial personal end (as in the original Phone Operator) as when he is acting for the sake of a weighty personal end (as in Phone Operator With Urgent Call). If this is correct, there is a presumptive case for thinking that each such threatener is liable to defensive harm. This case appeals to the stringency of the evidence-relative duty against risk imposition which the threatener is infringing: it invokes the moral significance of his responsible, risk-imposing agency, not the passive movement of his body.

To be sure, we may still be able to resist this conclusion. But we must do so, I believe, by demonstrating that the relevant threatener's evidence-relative duty against imposing the relevant risk of harm is *not* highly stringent. If that is so, he will not be obligated to internalize the risk, nor will he be liable to harmful defensive measures that in effect forces him to internalize it. But, to repeat, such a threatener is about to infringe a highly stringent fact-relative duty against killing another person.[169] And so, *presumptively,* his evidence-relative duty against imposing any given risk of this harm is highly stringent as well. What we need is some principled rationale for thinking that this presumption is undercut in the context at hand: we need to isolate some morally significant factor that substantially reduces the presumptive stringency of such a threatener's evidence-relative duty against risking harm to his victim.

---

[169] If an agent is threatening to infringe a fact-relative duty against harm that is not highly stringent, then his evidence-relative duty against imposing any given risk of this harm will not be highly stringent either. This is the case in some of the cases discussed in this article's predecessor, such as Unowned Vial and Threatening Organ Donor. See "Strict Liability to Defensive Harm I," pp. 31 – 32 and pp. 33 – 34. It might be true in other cases as well. For example, some philosophers believe that the fact-relative duty against harming another person by redirecting an existing threat towards him is much less stringent than the fact-relative duty against harming another person by subjecting him to a threat of one's own creation. Other philosophers believe that the fact-relative duty against harming another person, by (unintentionally) causing a responsible third party to harm him, is much less stringent than the fact-relative duty against harming another person by intruding on his body. These claims are controversial, but even if they are correct, neither will help us here. For in Phone Operator, Light Switcher, Falling Man, and other such cases, the relevant threatener is about to intrude onto another person's body, thereby infringing a stringent fact-relative duty against killing that person. And the point is not confined to these particular cases; it applies to indefinitely many situations involving astronomically small risks of harm.

But such a rationale is more difficult to come by than we might hope. The point can be demonstrated by closely examining the most familiar argument for exempting "non-responsible threats" like the Phone Operator from strict defensive liability: the idea that the harmful consequences of their actions are not *reasonably foreseeable* to them.

II. The Illusory Appeal of "Foreseeability"

Reflecting on Phone Operator With Urgent Call and Magical Button immediately casts doubt on this suggestion about foreseeability. For in that case, the possibility that his action might have harmful consequences is *actually foreseen* by the relevant threatener; it is precisely because he foresees this risk of harm that he must decide whether to internalize it. It is difficult to understand how a possibility of harm that is actually foreseen by a particular person can be unforeseeable to him, in any natural and non-fictive sense of the term.

The case is utterly fantastical, of course. In the real world, an agent who threatens to kill another person in some astronomically improbable way will rarely, if ever, have brought this possibility to mind before unleashing his threat. But the case just illustrates, in a particularly vivid manner, a point with more general application. As many cognitively mature adults know, and all such adults are in a position to know, every human action creates indefinitely many extremely small risks of harm. For this reason, even when an agent cannot foresee the particular, astronomically improbable harmful causal chain that his action will trigger, he *can* foresee that his action might trigger *some* such harmful causal chain. And it is the foreseeability of harm, not the foreseeability of any particular causal pathway to harm, that seems to ground both strict defensive liability and obligations to internalize risk. To see this, consider

> *Two Buttons:* Threatener A knows that, if he presses a red button, there is a 1% risk he will thereby cause a certain fighter jet which is flying around the world to drop a bomb, killing an innocent person. Threatener B knows that, if she presses a blue button, there is a 1% risk that she will kill another innocent person (in a way that infringes a stringent fact-relative duty against killing). However, Threatener B does not and indeed cannot foresee the way in which this death would come about. For a villain has given her a drug, seriously compromising her ordinary faculties for

imaginative foresight. Both threateners have equally weighty reasons, of the very same kind, for pressing their respective buttons. As an impartial Observer knows, both risks are about to eventuate. There appears to be no morally relevant difference between these two risky actions: it seems that Threatener A is obligated to internalize the 1% risk of death that he is creating just in case Threatener B is obligated to internalize the 1% risk of death that she is creating. So too, it seems, each threatener is liable to defensive harm just in case the other is. The impartial observer has no reason to prefer defensively harming one threatener over defensively harming the other.

If the point does not convince, we may next observe that even astronomically improbable causal chains *can* be foreseen, in their detailed particularity. A distinguished Australian jurist once remarked that "I cannot understand why any event which does happen is not foreseeable by a person of sufficient imagination and intelligence."[170] The point is even more forceful when a particular person does foresee the particular event in question, whether through imagination, intelligence, or mere luck. Suppose that our phone operator has just come across some articles on the liability of "non-responsible threats." He is thus fully aware that his phone might have been programmed by a villain, such that it will detonate a bomb and blow up his neighbor's house once it is dialed.[171] In a feat of cosmic coincidence, this very thing is going to happen when he dials his phone, unless an impartial observer defensively kills him beforehand. It seems quite clear that bringing this astronomically improbable possibility to mind cannot render this threatener liable to defensive killing, if he would otherwise have been exempt from it. But here our threatener foresees (and is, a fortiori, able to foresee) not only the possibility that his action will kill someone, but the very causal chain by which this harmful consequence will come about.

Perhaps the point should not be pressed too far. Reflective legal minds have long understood that all harms, however unlikely, are foreseeable in the sense of conceivable. Still the law has often insisted that sufficiently unlikely harms, or some of them, are unforeseeable in some different sense of the term. Perhaps

[170] *Chapman v. Hearse* (1961) 106 CLR 112, 115.
[171] For a similar thought experiment, see Ferzan, "Culpable Aggression," p. 680.

we should look to something like this legal sense of foreseeability (or one of them, if there are many)[172] in order to exempt threateners like the phone operator from strict defensive liability. I doubt, however, that any such legal sense of foreseeability will carve close to the moral joints. It may be tolerable and sensible for a legal system to operate with a conception of foreseeability according to which "what is 'foreseeable' is 'the kind of thing that happens'; something which we determine on the basis of a mishmash of personal experience, popular culture, and anecdote."[173] The pragmatic, incremental, institutionally circumscribed character of legal decision-making generates no strong pressure to clean the conceptual mishmash up. But the task is unavoidable for moral philosophers, since our charge is to uncover the content of the fundamental moral principles that govern this domain. And when we attempt the task, I think, we find no conception of foreseeability that is suitable to discharge it. No available conception of foreseeability seems equipped to explain why astronomically improbable risks of harm are unforeseeable (such that they need not be internalized, and do not generate liability to defensive harm).

Consider, for example, Susanne Burri's interesting suggestion that a harm is foreseeable to a threatening agent just in case she has a moral duty of care, of the kind recognized by the law of negligence, to guard against inflicting this harm.[174] Thus it is foreseeable that conscientious driving might kill a pedestrian, for drivers have a duty of care to guard against such possibilities (by keeping their eyes on the road, regularly taking their cars to the mechanic, and so on). By contrast, no one has a duty of care to guard against the possibility that he might detonate a bomb by dialing his phone, or electrocute his neighbor by flipping his light switch. These risks are simply too fantastical, too remote.

Suppose, however, that in a hundred years all cars have been replaced by self-driving cars, each of which poses a small risk of harm to pedestrians whenever it gets on the road - a risk of roughly equal magnitude to the risks created by conscientious driving in our own day and age. Suppose there is nothing

---

[172] On the different notions of foreseeability at work in tort and contract, see generally J.C.P. Goldberg and B. Zipursky, "Vosburg v. Baxendale: Recourse in Tort and Contract," in *Civil Wrongs and Justice in Private Law,* eds. P.B. Miller and J. Oberdiek (Oxford: Oxford University Press, 2020): 463 - 484.
[173] H. Scott, "The History of Foreseeability," *Current Legal Problems* 72 (2019) 1: 287 - 314, 313.
[174] S. Burri, "Morally Permissible Risk Imposition and Liability to Defensive Harm," *Law and Philosophy* 39 (2020) 4: 381 - 408.

anyone who uses a self-driving car can do to alter the magnitude of these risks. Then no such threatener will have a duty of care to guard against the harms that might result. But this would not seem to bear on whether any such threatener is liable to defensive harm, should the risk he is imposing on a pedestrian materialize. Whether a threatener can and should guard against a harm bears on whether he is at *fault* for inflicting it; that is why the duty of care inquiry is a lodestar of the law of negligence.[175] When a threatener *has* taken all due care against inflicting some harm, the fact that he had a duty of care against inflicting it seems entirely ill-suited to ground his liability.

It might be responded that the user of any such self-driving car would be obligated to guard against such risks to pedestrians, *if she could*. That is surely right. But it is equally true that we would all be obligated, *if we could*, to guard against the astronomically small risks of harm that our actions impose on others. If a person were presented with a magical button that would instantaneously eliminate all such risks, at no cost to herself or anyone else, she would plainly be obligated to press the button (*Comprehensive Risk-Eliminating Magical Button)*. To be sure, it might well be permissible for the person to refrain from pressing such a magical button if she would thereby impose equivalent risks on herself (*Comprehensive Risk-Internalizing Magical Button*). But that putative permission is precisely what we are trying to explain: why, if at all, would it be permissible for each of us to refrain from the internalizing astronomically small risks of harm we impose on other people, even if we could internalize these risks?

Appealing to duties of care, whether directly or in the conceptual guise of "foreseeability," will not help us in this task. The point can be reinforced by another fantastical example about a magical button. Suppose God decides to test Abraham's devotion to the moral law by providing him with a certain magical button. Every day, by pressing this button in the morning, Abraham can massively reduce (but not eliminate) the astronomically small risk that his ordinary activities will kill another person through some freakishly unlikely chain of events. If Abraham fails one day to press the button, he will arguably have failed in his duty of care. He may thereby incur fault-based liability to defensive harm, if any of the relevant risks

[175] See generally J.C.P. Goldberg, A.J. Sebok, and B. Zipursky, *Tort Law: Responsibilities and Redress* 4th ed. (New York: Wolters Kluwer, 2016): 47 – 143.

should threaten to materialize. But suppose that Abraham dutifully presses the button each morning. And suppose that he would have been exempt from strict defensive liability for the eventuation of astronomically improbable risks of harm if God had not provided him with the button. Surely he is not strictly liable for the eventuation of such risks *now*, simply because his acquisition of the magical button has saddled him with a duty of care which he is scrupulously discharging. Coming under a duty of care can create new opportunities for fault, but it cannot conjure up strict liability where none existed before.

I have not yet discussed the conception of foreseeability that may seem most suitable to our purpose. Perhaps a risk of harm is unforeseeable, in the relevant sense, just in case its magnitude falls beneath some specified quantitative threshold. The location of such a line would be exceedingly difficult to discern, of course, and it might even be metaphysically indeterminate.[176] But that sort of problem is ubiquitous in moral philosophy, and it does not strike me as distinctively problematic here.[177] There is a more damning problem, however, which is simply that it is intolerably arbitrary to draw any such line *at all* in the absence of a principled explanation for doing so. The task at hand is precisely to explain why there is any such line, any threshold beneath which risks of harm need not be internalized and do not generate defensive liability. Invoking the quantitative threshold conception of "foreseeability" is nothing more than bald assertion that the desired line exists.

There is another, even more damning problem with this conception of foreseeability. Let the relevant line be drawn at 1/x, where x is some sufficiently large number (say, ten trillion). Now compare two scenarios. In the first, each of x-1 threateners imposes a 1/(x-1) risk of death on Dave. This state of affairs is common knowledge among all the parties involved. As all of these parties know, therefore, it verges on a statistical certainty that one of the x-1 threateners will kill Dave. The quantitative threshold conception of foreseeability implies that it is foreseeable to each of the threateners that her action might kill Dave. Each

[176] Burri, "Morally Permissible Risk Imposition," pp. 391.

[177] It is, for example, exceedingly difficult to say just how many people a trolley must be threatening before one can permissibly throw a bystander into its path, or how many strangers' lives must be at stake before one is obligated to let one's child die in order to save them instead. In fact, it is possible there is no particular number at which such actions turn from determinately impermissible to determinately permissible.

threatener is liable, therefore, to be defensively killed if the risk she is imposing on Dave should threaten to materialize. That seems plausible enough. But now compare a second scenario, in which each of x+1 threateners imposes a 1/(x+1) risk of death on Dave. Here the quantitative threshold conception of foreseeability implies that it is *not* foreseeable to any of these threateners that her action might kill Dave - and that no threatener is liable to be defensively killed on his behalf.

As a conceptual matter, this seems like a strange conception of foreseeability, given that it verges on a statistical certainty (in both scenarios) that one of the threateners will end up killing Dave. But the more important point is that embracing this conception of foreseeability produces a bizarre conclusion about each threatener's liability to defensive harm. It is implausible to suppose that an impartial observer's permission to save Dave's life, by defensively killing the threatener who is about to kill him, should evaporate because two additional people have been added to the extremely large pool of threateners who are imposing risks on Dave - risks that (in aggregate) already make Dave's death a virtual certainty, even before these additional two threateners are added to the pool.

Perhaps there is some other conception of foreseeability that is capable of underwriting the desired conclusion, that we need not internalize astronomically small risks of harm (even if we can) and are not liable to defensive harm upon their eventuation. I doubt it, but the possibility cannot be ruled out without exhaustively cataloguing and refuting the candidate conceptions. Nevertheless, here is a general reason to be suspicious of the appeal to foreseeability in this context, however precisely this idea is ultimately understood. As I suggested in the predecessor to this article, obligations to internalize risk and strict defensive liability are plausibly regarded as forms of substantive moral responsibility for the consequences of risk imposition.[178] And it is plausible, as a general matter, to ground our substantive moral responsibility for various outcomes in their connection to our capacities for responsible agency. If the possibility of some event is incapable of entering into an agent's deliberation, therefore - if it is not foreseeable, in the ordinary and straightforward sense of the term - we may have grounds to exempt the relevant agent from substantive

[178] "Strict Liability to Defensive Harm I," p. 21.

moral responsibility for bringing this event about. But there will be such basis for exemption if the possibility of the relevant event *is* capable of entering into an agent's deliberation. And we have seen that, even if astronomically improbable harms are unforeseeable in some technical, legal, or otherwise non-literal sense of the term, they can indeed figure in a moral agent's responsible deliberation. My appeal to fantastical risk internalization devices such as magical buttons has been, in part, an attempt to illustrate this possibility in a sharp and vivid way.

Some may disdain the probative force of such hypothetical fancies. But the point can also be demonstrated, if somewhat less sharply, by considering more prosaic instruments for internalizing risk. The law of contract relieves a breaching promisor of any duty to pay damages, for the harmful consequences that his breach inflicts on his contractual counterparty, when those harms are (in some relevant, quasi-technical sense) unforeseeable.[179] But this exemption from contractual liability is only a default rule, which the contracting parties are free to alter by advance stipulation. By contracting around this rule, therefore, a promisor can commit to internalize whatever harmful consequences her breach might inflict on her counterparty, no matter how unforeseeable (in the relevant, quasi-technical sense) these consequences may be. Such a contract is a device by which an agent can internalize certain risks of avowedly unforeseeable harm that she might inflict on another person.

My point is not that the moral dynamics of strict defensive liability can be inferred from the moral dynamics of contractual risk internalization. My point is that, as the existence of such legal devices indicates, it is possible to internalize a risk of harm no matter how unforeseeable (in various non-literal, quasi-technical senses of the term) the harm in question may be. These risks can therefore enter into an agent's responsible deliberation, in the way that substantive moral responsibility for their eventuation arguably requires. If we wish to deny an agent's substantive moral responsibility for the eventuation of such risks, therefore - whether in the form of obligations to internalize these risks, strict defensive liability, or duties to compensate upon

[179] *Hadley v. Baxendale,* 9 Exch. 341 (1858); *Restatement (Second) of Contracts* §§ 351(1), (2)(b) (1979).

their eventuation - appealing to the putative unforeseeability of such risks is unlikely to help us. Some other rationale is required. And in the case of strict defensive liability, at least, the required rationale is elusive.[180]

As I have already suggested, I can think of only three rationales that seem promising. The first rationale is that astronomically small risk impositions are *reciprocal* in character. The second rationale involves the fact that merely existing in the world imposes a certain degree of risk on other people. The third rationale involves the fact that astronomically small risk impositions not only risk *harm* to other people, but confer astronomically small expected *benefits* on other people as well. We might hope that one of these considerations will relieve us of the obligation to internalize astronomically small risks of harm, and preclude us from incurring strict defensive liability for the eventuation of such risks. But none of these considerations, I will argue over the next several sections, can do so as comprehensively as we might wish.

III. The Moral Significance of Reciprocal Risk Imposition

The idea that reciprocal risk imposition bears on obligations to internalize risk, and strict defensive liability, was discussed at several points in this article's predecessor. Recall

> *Reciprocal Russian Roulette:* In order to save his life from a near certainty of death, Tomas must impose a 1% risk of death on Victoria, by shooting at her a revolver which contains a hundred chambers (one of which is loaded). Tomas has no way to internalize this risk (as Victoria is aware). Meanwhile, in order to save *her* life from a near certainty of death, Victoria must impose a 1% risk of death on Tomas, by shooting just such a revolver at him. Victoria *can* internalize this risk, by enabling a feature on her gun that will blow it up (killing her) if and only if it is about to trigger the loaded chamber (so killing Tomas).[181]

It seems to me intuitively that Victoria is not obligated to internalize the risk of death that she is imposing on Tomas, given that (as she knows) Tomas is not internalizing the equivalent risk that he is imposing on

[180] In the case of contractual obligation, the rationale may be easier to come by. Exemption from responsibility for "unforeseeable" (i.e. sufficiently unlikely or qualitatively unusual) consequences of breach may be something like an implicit condition of many contractual arrangements. See e.g. M.E. Eisenberg, "Impossibility, Impracticability, and Frustration," *The Journal of Legal Analysis* 1 (2009) 1: 207 - 261.
[181] "Strict Liability to Defensive Harm I," p. 42.

her. If this is correct, the risk internalization principle implies, then Victoria will not be liable to defensive killing in order to prevent the risk she is imposing on Tomas from materializing. Nor is Tomas liable to be defensively killed in order to prevent the risk that he is imposing from Victoria from materializing. Faced with a tragic situation of this kind, an impartial observer should simply refrain from intervention.

These conclusions strike me as intuitively plausible. Similarly, I have argued, it is plausible that a conscientious driver whose car threatens to kill another person is not liable to be defensively killed, if that person is currently imposing an equivalent risk of death on the driver. That is true in Conscientious Driver Threatening Another Conscientious Driver.[182] So, too, if the other person has previously imposed an equivalent risk on the driver. That is true in Conscientious Driver Threatening Car-Owning Pedestrian.[183] In neither case, I suggested, would the relevant threatener be obligated to internalize the risk that he is imposing on his victim, even if he could.

If my claims about these various cases are plausible, we may have a principled basis for exempting threateners who impose astronomically small risks of harm on other people from their presumptive obligations to internalize these risks, and their presumptive liability to defensive measures that will in effect force them to internalize these risks. Each of us knows, or should know, that each of her actions may result in a threat of astronomically improbable harm to another person. In that sense, we are all unavoidably and ubiquitously in the business of imposing astronomically small risks on one another. And it is reasonably plausible to suppose that all of the relevant risks are equally small - even if (or partly because) the exact magnitude of each such risk is indeterminate.[184] In this respect, the most quotidian spheres of ordinary human activity may be like a highway populated by perfectly conscientious drivers, each of whom is imposing the same set of risks on the others. If no driver is obligated to internalize the risks she is imposing on her fellows - and no driver is liable to defensive harm should such a risk materialize - perhaps the imposition of astronomically small risks in ordinary life can be treated on the same model.

---

[182] *Id.* at 64.
[183] *Id.* at 63.
[184] On closer inspection, there are reasons to doubt that all such risks are indeed equally small. See below, pp. 123 - 124.

Invoking reciprocal risk imposition is appealing, therefore, partly because it might secure these intuitively plausible results about (the absence of) strict defensive liability for car driving and astronomically small risk imposition. But it also has more direct case-based appeal, as illustrated by cases like Reciprocal Russian Roulette. There are many reasons, therefore, that such a view warrants close examination and careful development. Here is an initial pass at formulating it:

> THE RECIPROCAL RISK ACCOUNT: Suppose an agent A's evidence-relative duty against imposing some risk of harm on another person B is sufficiently stringent that she is presumptively obligated to internalize this risk (if she can) and liable to defensive measures that will in effect force her to internalize this risk (even if she cannot). And suppose that B is imposing (or did previously impose) a comparable risk of harm on A, which he is not internalizing (or did not internalize). Then the ordinary stringency of A's evidence-relative duty against imposing the relevant risk is substantially undercut, such that she is permitted to refrain from internalizing this risk (even if she can) and is not liable to defensive measures that will in effect force her to internalize it.

So formulated, the reciprocal risk account is not a comprehensive account of strict defensive liability.[185] It is merely an account of one significant factor that reduces the ordinary stringency of evidence-relative duties against risk imposition - thereby undercutting obligations to internalize risk, and strict defensive liability, that would otherwise attend a risk-imposing act. Other factors (such as the fact that a threatener is acting for the benefit of another person) may well perform a similar function.[186]

According reciprocity this general sort of moral significance - the ability to alter, weaken, or even eliminate the obligations and liabilities that interacting parties would otherwise have in respect of one another - seems to be a deeply rooted feature of human social life, one that appears in a variety of legal,

---

[185] For this reason, I believe, it evades the nice criticisms that Adam Slavny presses against reciprocal risk accounts of corrective liability in tort law (which are typically more ambitious). See A. Slavny, "Nonreciprocity and the Moral Basis of Liability to Compensate," *Oxford Journal of Legal Studies* 34 (2014) 3: 417 - 422. For a compelling reciprocal risk-based account of tort law that may also evade these criticisms, see A. Stein, "The Domain of Torts," *Columbia Law Review* 117 (2017): 535 - 611.

[186] "Strict Liability to Defensive Harm I," pp. 73 – 80.

political and cultural contexts.[187] A nice illustration, if a somewhat obscure one, is provided by the private law doctrine of set-off.[188] Suppose I have wronged you and thus owe you a sum of money. Meanwhile you have wronged me and thus owe me an equivalent sum. If you are unable to pay your debt to me because you are insolvent, I am absolved of my duty to pay you: the law regards it as "natural justice and equity, that in all cases of mutual credit, only the balance should be paid…"[189] The impulse seems morally sound. Even if considerations of reciprocity cannot *ground* the relevant duties and liabilities,[190] it seems plausible they can undercut duties and liabilities that arise on independent grounds.

Given this congruence with compelling general patterns of moral thought, and its ostensible explanatory power, it is surprising that reciprocal risk-based approaches to strict defensive liability have not been developed or examined at any length. Reciprocal risk-based approaches to corrective liability in tort law *have* been developed at length, first in a famous article by George Fletcher and later in important work by legal scholars such Gregory Keating, Mark Geistfeld, and Alex Stein.[191] We might expect this literature to provide us with a ready template for thinking about reciprocal risk in the context of defense, for it is plausible to suppose that defensive liability and corrective liability are closely connected.[192] But the template is imperfect, since this literature has a largely interpretive and political cast: it seek to capture the doctrinal contours of Anglo-American tort law, and frequently appeals to considerations of distinctively political or institutional morality (often Rawlsian in character) in order to do so.[193] Similarly, in critical responses to this

---

[187] See M.A. Geistfeld, "Hidden in Plain Sight: The Normative Source of Modern Tort Law," *New York University Law Review* 91 (2016): 1517 - 1594.
[188] See generally W.H. Loyd, "The Development of Set-Off," 64 *University of Pennsylvania Law Review* 541 (1916).
[189] *Lord Lanesborough v. Jones,* 1 P. Wms. 325, 24 Eng. Rep. 409 (Ch. 1716).
[190] For the view that it can, see D. Rodin, "The Reciprocity Theory of Rights," *Law and Philosophy* 33 (2014) 3: 281 - 303.
[191] G.P. Fletcher, "Fairness and Utility in Tort Theory," *Harvard Law Review* 85 (1972): 537 - 73; G. Keating, "A Social Contract Conception of the Law of Accidents," in *Philosophy and the Law of Torts,* ed. G. Postema (Cambridge: Cambridge University Press, 2001); Geistfeld, "Hidden in Plain Sight"; Stein, "The Domain of Torts."
[192] S. Bazargan, "Defensive Liability Without Culpability," in *The Ethics of Self-Defense,* eds. C. Coons and M. Weber (Oxford: Oxford University Press, 2016): 69 - 86; McMahan, "Basis of Moral Liability," p. 395; V. Tadros, "Causation, Culpability, and Liability," in *The Ethics of Self-Defense,* eds. C. Coons and M. Weber (Oxford: Oxford University Press, 2016): 111 - 130, 118.
[193] Fletcher, for example, argues that tort liability is explained by considerations of Rawlsian political morality, according to which each inhabitant of a political society has a right to equal security that is violated when she suffers harm through the non-reciprocal imposition of a risk upon her.

literature, objections from interpretive adequacy have loomed large.[194] To be sure, despite its largely interpretive and institutional character, this literature also contains many valuable insights about the interpersonal morality of imposing non-reciprocal risks. But since our own concerns lie exclusively at the level of fundamental moral principle - and since defensive and corrective liability may, despite their significant connections, differ in important ways - it may be wise to conduct our investigation on a relatively clean slate.[195]

Let us proceed, then, with the enterprise of developing the reciprocal risk account of strict defensive liability - before considering whether it can exempt astronomically small risk imposition from incurring defensive liability, as we might hope. I will mostly discuss obligations to internalize risk, rather than strict defensive liability, but in each case the two phenomena should be co-extensive.

First, we need to specify just when two risks of harm are "comparable" in the relevant sense. It is tempting to do so exclusively by reference to the magnitude of the harm that is risked and the magnitude of the risk itself. Call this the *simple magnitude* approach. It yields sensible results in certain cases, but it misfires in others. Consider

> *Roulette and Vial:* In order to save himself from a very large risk of death, Tomas imposes a 5% risk of death on Victoria by playing Russian Roulette on her. In order to save herself from a very large risk of death, Victoria imposes a 5% risk of death on Tomas by treating herself with an unowned vial of medicine that will probably (but not certainly) save her, and which Tomas may also need. Both actions are evidence-relative permissible (although each may be, depending on the facts, fact-relative impermissible). Each party knows about the other's action.

Tomas's evidence-relative duty against imposing a 5% risk of death on Victoria, by shooting at her, is very stringent: he is presumptively obligated to internalize this risk. Victoria's evidence-relative duty against imposing a 5% risk of death on Tomas, by using up the unowned vial, is not nearly as stringent: she is *not*

[194] See e.g. J. Coleman, *Risks and Wrongs* (Oxford: Oxford University Press, 2002): 253 - 269.
[195] Insofar as the reciprocal risk accounts of strict corrective liability in the tort literature are concerned with fundamental moral principle rather than legal interpretation or institutional design, the difficulties we will encounter in the domain of defensive liability will arise for these accounts as well.

presumptively obligated to internalize this risk.[196] The simple magnitude approach implies that these risks are suitably reciprocal, such that the ordinary stringency of Tomas's evidence-relative duty against imposing the relevant risk is substantially undercut. But that seems wrong: it seems that Tomas is still obligated to internalize the risk he is imposing on Victoria, while Victoria is not obligated to internalize the risk she is imposing on Tomas. This suggests that we should reject the simple magnitude approach. Instead we should understand "comparability" as a relationship between the stringency of the different evidence-relative duties involved. This will be a function both of the magnitude of the relevant risks and of the deontological moral character of the harms which are being risked.

Second, we need to figure out how to deal with risks that are nearly equivalent in magnitude but not exactly so. Consider

> *Two Buttons:* In order to save himself from a near certainty of death, Tomas is about to press a red button, which he knows will impose a 1% risk of death on Victoria. Tomas cannot internalize this risk. Meanwhile, in order to save himself from a near certainty of death, Victoria is about to press a blue button, which will impose a 1.01% risk of death on Tomas. Victoria *can* internalize this risk. Each party knows about the other's action.

Must Victoria internalize the 1.01% risk she is imposing on Tomas, while leaving herself open to the 1% risk that Tomas is imposing on him? No, I think, given that the additional quantum of risk (0.01%) that Victoria is imposing on Tomas is very small. Matters would be different if Victoria could internalize *only* this additional quantum of risk.[197] But here she must choose whether to internalize the entirety of the risk she is imposing on Tomas or to internalize none of it. I do not think she is obligated to internalize it all.

---

[196] See the discussion of Unowned Vial in "Strict Liability to Defensive Harm," pp. 31 – 32, 34 – 35.

[197] As in the following case:

> *Three Buttons:* In order to save himself from a near certainty of death, Tomas is about to press a red button, which he knows will impose a 1% risk of death on Victoria. Tomas cannot internalize this risk. Meanwhile, in order to save himself from a near certainty of death, Victoria is about to press a blue button, which will impose a 1% risk of death on Tomas, and a green button, which will impose a 0.01% risk of death on Tomas. She can internalize either risk, both, or neither.

Victoria is obligated to internalize the 0.01% risk that, by pressing the green button, she is imposing on Tomas.

At what point, though, does the additional quantum of risk become so large that the moral situation changes? Consider

> *Two Buttons*:* In order to save himself from a near certainty of death, Tomas is about to press a red button, which he knows will impose a 1% risk of death on Victoria. Tomas cannot internalize this risk. Meanwhile, in order to save herself from a near certainty of death, Victoria is about to press a blue button, which will impose a 2.5% risk of death on Tomas. Victoria *can* internalize this risk. Each party knows about the other's action.

Here I am inclined to think that Victoria must internalize all of the 2.5% risk she is imposing on Tomas rather than internalize none of it. But where is the line to be drawn, and what explains where it lies? Perhaps there is no determinate answer. That would be disturbing, but there are all sorts of similar line-drawing problems in ethics; the problem may be a very general one. Perhaps there is nothing more to say.

But perhaps we do have something helpful to say. In the predecessor to this article, I suggested that there may be a distinctively deontological conception of fairness, according to which it is unfair that one person should take a greater gamble with another person's rights than the gamble which she is taking with his.[198] It may be the distinctively deontological unfairness of such a relationship, I suggested, that explains why an agent need not internalize a risk she is imposing on another person when that other person is imposing an equivalent risk on her. This suggests a promising approach to our present question. In both Two Buttons and Two Buttons*****, Victoria must choose between (a) taking a greater gamble with Tomas's rights than he is taking with hers and (b) allowing Tomas to take a greater gamble with her rights than she is taking with his. Perhaps Victoria may permissibly elect the former option only if it is less unfair to Tomas than the latter option is unfair to her. If that is correct, then Victoria may impose without internalizing up to a 2% risk of death on Tomas, but she must internalize any risk that falls above this threshold. (Again, all of this assumes that Victoria is deciding between internalizing and externalizing the entirety of the relevant

[198] "Strict Liability to Defensive Harm I," pp. 42- 43.

risk. If she can internalize *only* the additional quantum of risk of that she is imposing on Tomas, then she is obligated do so.)

The suggestion is not obviously correct, but nor is a better approach not obviously available. So we are already seeing an unsettling broader point come into view. The reciprocal risk account is quite compelling in the abstract. And it is very compelling as applied to simple cases like Reciprocal Russian Roulette, in which the relevant risks and modes of harm are exactly the same. But it immediately loses some of its appeal when we introduce even a modicum of additional complexity into our cases.

Third, we need to deal with situations involving more than two parties, such as the following cases (in which, let us suppose, all of the parties are imposing their risks justifiably, and for the sake of equally weighty self-regarding ends):

> *Triangular Russian Roulette:* Alberto is imposing a 1% risk of death, via Russian Roulette, on Barney; Barney is imposing a 1% risk of death, via Russian Roulette, on Cornelius; Cornelius is imposing a 1% risk of death, via Russian Roulette, on Alberto. Alberto cannot internalize the 1% risk he is imposing on Barney, and Barney cannot internalize the 1% risk of death he is imposing on Cornelius. However, Cornelius *can* internalize the 1% risk of death he is imposing on Alberto.[199]

It does not seem to me that Cornelius is obligated to internalize the 1% risk he is imposing on Alberto, while leaving himself open to the 1% risk of death Barney is imposing on him. But strictly speaking, none of the relevant risks are reciprocally imposed: no one is imposing a risk on a person who is imposing a risk on *him*. So the reciprocal risk account, as we have formulated it, provides no ground for the relevant exemption.

Perhaps, however, a principled reformulation is available. In the original Reciprocal Russian Roulette, one way to explain why Victoria need not internalize the 1% risk she is imposing on Tomas is by reference to a hypothetical contract that the parties would be obligated to strike, if they could. Under this contract, each party would waive her right against being risked death by the other party, in exchange for

[199] Compare the structure of the central case in R. Christopher, "Self-Defense and Defense of Others," *Philosophy & Public Affairs* 27 (1998) 2: 123 - 141.

the other party's waiver of the corresponding right against being risked death by her.[200] If this basic line of reasoning is sound, a modest extension of it might explain Triangular Russian Roulette: if they could communicate, each party would be obligated to enter an agreement with the others, whereby each party would waive his right against being risked death by one of his counterparties and receive in exchange the waiver of his *other* counterparty's right against being risked death.[201] Alternatively, we might try to explain the situation by directly appealing to the unfairness of a situation in which Cornelius is subjected to a 1% risk of death without being allowed to subject Alberto to a 1% risk of death in turn. While neither strategy is obviously successful, each has some promise and may warrant further development. It is worrying, however, that both strategies seem to move away from the compelling underlying idea with which we started: the idea of *reciprocity,* that our claims against others are sensitive to whether we honor *their* claims on us. Some distance from this idea is compelled by the very structure of these multi-party cases.

Fourth, we need to formulate the reciprocal risk account so that it can handle cases involving very large numbers of risk impositions, such as the following:

> *1001 Risk Impositions:* Each of a thousand people independently and justifiably imposes a 1% risk of death on Sally (in each case in order to save himself from a near certainty of death). Miraculously, none of these risks materialize; Sally emerges safe and sound. Now, however, Sally's own life comes under threat and she can save it only by imposing a 1% risk of death on each of the thousand.

---

[200] Compare the structure of the arguments in K. Draper, *War and Individual Rights* (Oxford: Oxford University Press, 2015): 237 - 250; V. Tadros, *The Ends of Harm* (Oxford: Oxford University Press, 2011): 274 - 275.

[201] This line of reasoning *might* be able to explain a different, and even more challenging, kind of multi-party case:

> *Linear Russian Roulette:* Alberto is justifiably imposing a 1% risk of death, via Russian Roulette, on Barney; Barney is justifiably imposing a 1% risk of death, via Russian Roulette, on Cornelius. Barney cannot internalize the risk of death he is imposing on Cornelius. Alberto *can* internalize the risk of death he is imposing on Barney.

Is Alberto obligated to internalize the risk he is imposing on Barney? Intuitively, I am not sure. Maybe hypothetical contract reasoning can provide some guidance. Someone, either Alberto or Barney, is going to risk another person's death without being subjected to a risk of death in turn. Arguably, if the parties were to enter an agreement, they should agree to give Alberto and Barney an equal chance of this windfall benefit. (Perhaps Alberto should demand this of Barney, in exchange for waiving her own right against being risked death by him.) If so, then perhaps Alberto is permitted to flip a coin and, if he wins the flip, impose the relevant risk on Barney rather than internalize it.

It verges on a statistical certainty that by imposing these risks Sally will be killing ten of these people. And it is difficult to accept that she may do this, even in order to save her life. But because each person has imposed just such a risk on her, the reciprocal account implies that the ordinary stringency of Sally's duty against imposing the relevant risk on him is substantially undercut - such that she need not internalize this risk. Sally is just treating each of the one hundred in exactly the same way that he treated her, and so none of them is wronged, and so her action is not wrong, although it will almost certainly kill ten innocent people. That is counterintuitive.

We might try to avoid this result by maintaining that Sally's action would be wrong because of the impersonal disvalue of the ten deaths she would be causing, rather than because of her duties to any of the thousand. But in other contexts, it does not seem that an agent is required to let herself die in order to prevent the impersonal disvalue of ten deaths from coming about. (An agent would not be obligated to refrain from driving herself to the hospital for life-saving treatment in order to stop and save ten strangers from some threat to their lives.) Of course, in this case Sally would be *bringing about* the ten deaths, which is a different matter from allowing ten deaths to occur. In general, however, it is plausible that the moral significance of killing is principally or entirely explained by the weight of *claims* against being killed (or directed duties against killing). Bringing about impersonal disvalue does not seem much different, morally speaking, from allowing impersonal disvalue to occur.[202]

Another strategy for explaining why Sally may not kill ten people, by imposing a 1% risk of death on each of the thousand, is to maintain that the stringency of duties against risk imposition can *aggregate.*[203] Let us consider some arbitrary person (call him Ted) from among the one thousand. Ordinarily, Sally's duty against imposing a 1% risk of death on Ted is sufficiently stringent that she would be obligated to bear a 1% risk of death rather than infringe this duty. Now that Ted has imposed such a risk on her, however, the stringency of her duty against imposing such a risk on him is reduced. Still, we can plausibly maintain, this

[202] It is perfectly permissible to destroy one beautiful painting in order to save another that is only slightly more beautiful; by contrast, it is not permissible to inflict excruciating pain on one person in order to save another from suffering slightly worse pain.
[203] Some nonconsequentialists are in principle hostile to aggregation, but I am not one of them.

duty retains some considerable stringency. Thus Sally is obligated to bear up to (say) a 0.25% risk of death rather than infringe this duty. And if we aggregate the stringency of this duty to Ted with the stringency of Sally's *other* such duties (to the other nine hundred and ninety-nine people involved), we will obtain the result that Sally is obligated to bear up to a 250% risk of death in order to refrain from infringing all of these duties. So she must let herself die rather than infringe all these duties.

This may be the more promising strategy for handling our issue, but it too must answer some difficult questions. In particular, what is the discount function between the ordinary stringency of an agent's duty against imposing a risk on another person and this duty's stringency after she is subjected to a reciprocal risk by that person? Specifying the discount function with any precision is unlikely, of course, but we might at least hope for some principled guidance. But it is hard to see what such guidance would look like. Once again, then, the exercise of equipping the reciprocal risk account to handle more complex cases chips away at our conviction (my conviction, at least) in its underlying structure. It starts to feel less like we are elaborating the internal logic of a structurally sound basic moral idea, and more like we are making unprincipled adjustments in response to problem cases that put the basic idea under strain. The worry is of course a bit abstract and diffuse. The next worry we will consider is somewhat more concrete.

Fifth, an adequate specification of the reciprocal risk account will need to deal with problems involving the *identity* of the parties whose interests are risked. Consider

> *The Masked Man:* Sally has excellent reason to believe that her neighbor Dexter, an astronaut, is currently on a spaceship to the moon: it is incredibly unlikely (though possible) that he is in fact still on earth. Now a villain credibly threatens to murder Sally unless she imposes a 1% risk of death on a masked man by playing Russian roulette on him. Since Sally has excellent reason to believe that Dexter is on a spaceship to the moon, she has excellent reason to believe that the masked man is not Dexter. But the masked man *is* Dexter, who did not get on the spaceship because of some exceedingly unlikely turn of events. At the same time, Dexter is faced with a similar threat to his life, which he can avoid either by imposing a 1% risk of death on Sally or by imposing an equivalent risk of death on himself.

There are various descriptions (e.g. "my neighbor Dexter, the astronaut") under which Sally is imposing exceedingly little risk of harm on Dexter by playing Russian roulette on the masked man. But of course, there are also various descriptions (e.g. "this masked man in front me") under which Sally is imposing a 1% risk of death on Dexter. Which description is relevant for determining the magnitude of risk that Sally is imposing on Dexter, such that he may (under the reciprocal risk account) impose rather than internalize an equivalent risk on her? More generally, what determines the relevant description to which any given risk imposition is to be indexed? I have no general answer to this question. There are closely related issues concerning the morality of impartially distributing risks and harms across persons, and these issues have proved difficult to resolve.[204] The issue at hand for the reciprocal risk account may prove no more tractable.

Sixth, the reciprocal risk account will need to deal with cases involving *multiple* and *successive* risk impositions. Suppose Wanda saves her life from some threat by imposing a 1% risk of death on Vijay, which he survives; immediately thereafter she saves her life from a different threat by imposing another 1% risk of death on Vijay, which he again survives. Now Vijay can save himself from some threat to *his* life by imposing a 2% risk of death on Wanda or instead imposing a 2% risk of death on himself. The logic of the reciprocal account would seem to imply that Vijay may permissibly impose the risk on Wanda rather than internalizing it. For it seems irrelevant that Wanda has imposed a 2% risk of death on Vijay in the form of two successive 1% risk impositions, as opposed to a single lump sum. Similarly it seems irrelevant that Wanda has imposed this total amount of risk on Vijay *before* his choice to impose risk on her. So Vijay need not internalize the 2% risk of death he is imposing on Wanda, any more than if she were currently imposing a single 2% risk of death on him.

That seems like a plausible result. But once we are in the business of aggregating risk impositions across time, it seems that an agent's risk internalization obligations and strict defensive liability are going to be determined by (some suitable function over) a running tally of all the risks he has ever imposed, and to which he has been subjected, over the course of his entire life. And that is not an entirely comfortable place

[204] See especially A. Mahtani, "The *Ex Ante* Pareto Principle," *The Journal of Philosophy* 114 (2017) 6: 303 - 323.

to end up. Suppose that two retired military men, Ricardo and Simeon, find themselves in a tragic situation in which each needs to play 1/100 Russian roulette on the other in order to avoid some threat to his life. When the situation is viewed in isolation, the reciprocal risk account implies that neither is obligated to internalize the risk he is imposing on the other, and neither is liable to defensive harm. That is plausible. But on the line of thought we are now considering, the situation *cannot* be viewed in isolation; instead we need to examine the parties' life histories and tally up whatever risks they may have imposed on each other in their previous affairs. Suppose it turns out that thirty years ago, in the heat of some terrible military conflict, Ricardo justifiably imposed a significant risk of death on Simeon in order to escape some threat to his own life. If this information comes to his attention, Ricardo will find himself evidence-relative obligated to internalize the 1% risk of death he is imposing on Simeon while leaving himself open to the 1% risk of death that Simeon is imposing on him.

It is at least somewhat strange to think that the distant past should cast so powerful a shadow over the obligations and liabilities that our threatening activities incur in the present. But once we abandon any artificial stricture that reciprocal risk impositions must be simultaneous, and once we allow that small risks of harm may aggregate so as to offset the reciprocal imposition of a single larger risk of harm, this conclusion seems inexorably to follow. Each agent's life will be like a ledger in which all the risks she has ever borne or imposed are permanently inscribed, and in some cases an agent's present obligations and liabilities will be determined almost entirely by this standing ledger, and hardly at all by the moral character of the risky interactions in which she is presently involved. We might call this the *ledger feature* of the reciprocal risk account.

I have discussed these issues at some length in part because they bear on whether the reciprocal risk account ought ultimately to be accepted. Like many other ideas in deontological ethics - such as, for example, the idea that there is a moral asymmetry between doing and allowing harm - the reciprocal risk account is inherently plausible on initial inspection, and it promises to explain the compelling intuitive judgments about simple cases (such as Reciprocal Russian Roulette) that lead us to entertain it. But on closer inspection the idea admits of multiple candidate specifications, none of which is resonantly compelling, and

all of which may saddle us with some strange results. These difficulties should not lead us to hastily abandon the reciprocal risk account, any more than we should hastily abandon the moral asymmetry between doing and harm. But they should diminish our credence in it. Insofar as endorsing the reciprocal risk account proves necessary to explain other propositions - such as the idea that threateners need not internalize astronomically small risks, and are not defensively liable for their eventuation - we should reduce our credence in those propositions as well.

In any event, the reciprocal risk account *cannot* establish a categorical exemption from strict defensive liability for threateners who impose only astronomically small risks of harm. At least some threateners will fall outside the scope of any such exception, for at least three reasons.

The first picks up on our discussion of the reciprocal risk account's ledger feature. Consider

> *Military Veteran Phone Operator:* Some time ago, Threatener justifiably saved his life by imposing a large risk of death on another innocent person, Victim, in the heat of military conflict. Now that conflict lies in the distant past and both parties are leading peaceful and uneventful lives. Threatener is about to dial his phone to order some food for delivery. In an astronomically unlikely turn of events, a villain has programmed his phone such that it will trigger a bomb, destroying Victim's house and killing her, when he dials it. An impartial Observer knows this and can save Victim only by defensively shooting and killing Threatener with a sniper rifle.

If the situation is viewed in isolation, then Threatener would not be obligated to internalize the astronomically small risk of harm he is imposing on Victim (even if he could), and he is not liable to be defensively killed. Just as Threatener is imposing astronomically small risks of such harm on Victim, she is imposing small risks of such harm on him, and so the presumptive stringency of Threatener's duties against imposing these risks is substantially reduced. But if the ledger feature of the reciprocal risk account is sound, then we *cannot* view the situation in isolation. Since Threatener has previously imposed a large amount of risk on Victim, which has not been (and is not being) offset by any comparable risk imposition on him, his presumptive obligation to internalize the astronomically small risk of death he is imposing on her remains

in force. So too does his presumptive liability to defensive harm that will, in effect, force him to internalize this risk.

The preceding case is just an extreme instance, moreover, of a situation type that might arise with some frequency. Imagine a variant of Phone Operator in which Threatener drove his car past Victim's house yesterday, thereby imposing a very small risk of death on her. Then the overall amount of risk to which Threatener has subjected Victim might well be greater than the overall amount of risk to which Victim has subjected him. If so, then the reciprocal risk account will again supply no basis for exempting Threatener from defensive liability upon the eventuation of the astronomically small risk which he is imposing on Victim by dialing his phone.

The second reason derives from the fact that some astronomically small risk impositions are larger than others. Consider

> *Walkers in a Grove:* Threatener and Victim are strangers, walking in a grove. In an astronomically improbable turn of events, a freak gust of wind picks up Threatener and hurls him at Victim. Now Victim will be crushed by Threatener unless an impartial Observer blows him up with a rocket launcher. Threatener is an extremely large and heavy man whereas Victim is an extremely small and light woman.

By walking in the grove, Threatener and Victim are imposing astronomically small risks on each other. For example, Victim is imposing on Threatener the risk that *she* will be picked up by a freak gust of wind and hurled towards him with sufficient speed and force that she will kill him when she hits him. In light of their different sizes and weights, however, it is less likely that a freak gust of wind will produce this effect than the opposite effect - that is, Threatener killing Victim, the effect which is now going to obtain unless Observer defensively kills Threatener. In the relevant respect, Threatener and Victim are relevantly analogous to two drivers on the highway, one of whom is driving a bulky pickup truck and the other of whom is driving a compact sedan: one party is imposing greater risks on the other. The reciprocal risk account thus provides no basis for undercutting Threatener's presumptive obligation to internalize the risk he is imposing on Victim, or his strict liability to defensive measures that will effectively make him internalize this risk.

The third reason derives from asymmetries in cognitive and epistemic capacity between typical adult agents and people with profound cognitively limitations. Consider these cases:

> *Phone Operator and Cognitively Impaired Victim:* As in Phone Operator. However, Victim is a profoundly cognitively impaired adult, and thus incapable of grasping the fact that any of her actions might result in harm to other people.
>
> *Phone Operator and Little Baby Child:* As in Phone Operator. However, Victim is an infant, and thus incapable of grasping the fact that any of her actions might result in harm to other people. Plausibly, in fact, she does engage in action at all so much as instinctive bodily movement.

To impose a risk of harm on another person, in the relevant sense, is to possess evidence that one's action might result in such harm. In order to possess such evidence, however, it seems that an agent must have certain minimum cognitive capabilities. Otherwise she will be relevantly analogous to the lower animals, who cannot be said to have evidence of their actions' effects, insofar as they can be said to act at all. Neither of the victims above would seem to meet the requisite threshold of cognitive capacity. In neither case, therefore, can it be said that the relevant victim is imposing risks on Threatener. In these cases, the reciprocal risk account supplies no basis for undercutting Threatener's presumptive obligation to internalize the risks he is imposing on the relevant victim, or his presumptive liability to defensive harm.[205]

So the reciprocal risk account cannot support a categorical exemption from strict defensive liability for threateners who are imposing only astronomically small risks of harm. Still, arguably it *does* exempt some or even many such threateners from defensive liability. That is just one of several reasons to take the view seriously and seek to further develop it. Let me conclude this section by offering another. In a universe with a certain kind of architecture, of the kind contemplated by e.g. chaos theory, there is a non-trivial risk that even the most quotidian actions might kill another person - just as, in such a universe, "the flap of a

[205] The force of the point may be limited by the fact that some philosophers deny that babies and the profoundly cognitively impaired have as much moral status, or suffer as much harm by dying, as typical adults. If some such view is right, then any such victim may have weaker fact-relative claims against being killed, and weaker evidence-relative claims against being subjected to a risk of death, than a typical adult. Suppose for the sake of argument that some such view is right. Still it is plausible that duties against (risking) killing such victims have *considerable* stringency, such that Threatener will be strictly liable in each case to a significant amount of defensive harm, even if he is not liable to be killed.

butterfly's wings in Brazil [might] set off a tornado in Texas."[206] Our own universe may well possess such an architecture. If we acquire sufficient evidence of this fact, then our most quotidian actions - such as flipping a light switch or dialing a phone or going for a walk in a grove - will no longer pose only astronomically small risks of harm to others. They will pose decidedly non-trivial risks of harm - risks that responsible moral deliberation and moral theory will then need to grapple with, as a matter of urgent practical concern.[207] Even if we were to acquire such knowledge, however, it would still be intuitively tempting to exempt an agent who threatens to kill another person by (say) dialing her phone from strict liability to defensive killing. Or so it seems to me.

The reciprocal risk account seems like a promising strategy for explaining such intuitions: if we knowingly inhabited such a universe, we would all be imposing such non-trivial risks of harm on each another all of the time. Or at least, those of us with normal cognitive capabilities, and equipped with (or able to access) the relevant information, would be imposing such risks on each other all of the time.[208] The inhabitants of isolated civilizations, who did not possess the relevant information regarding chaos theory and so on, would *not* be imposing such risks on other people. Appealing to the reciprocal risk account would thus fail to block the conclusion that a scientifically informed Mexican would be strictly liable to defensive killing if he should threaten, by dialing his phone or walking in a grove, to kill the inhabitant of an isolated civilization - a consequence of the sort made readily foreseeable to him by his cognizance of his universe's causal architecture.

---

[206] E. Lorenz, "Predictability; Does the Flap of a Butterfly's Wings in Brazil Set Off a Tornado in Texas?" *American Association for the Advancement of Science,* 139th Meeting, quoted in H. Nye, "Chaos and Constraints," in *Dimensions of Moral Agency* ed. D. Boersma (Newcastle Upon Tyne: Cambridge Scholars Publishing, 2014): 24 n. 1.

[207] The significance of chaos theory is, even now, a matter of great concern for moral philosophy. The pathbreaking discussion is Nye, "Chaos and Constraints."

[208] In fact, it may be the case that those of us with normal cognitive capabilities and (access to) the relevant information are *already* imposing such non-trivial risks on each other all the time. For each of us should have some credence that we inhabit a universe with the necessary sort of causal architecture. Whether the relevant risks are properly classified as "non-trivial," rather than astronomically small, depends on how large our credence that we inhabit such a universe should be.

IV. The Baseline of Risk Imposed by Mere Existence

Merely existing in the world, as an entirely passive being, increases the likelihood that other people will suffer harm. The body of a newborn baby might be picked up by a gust of wind and flung at an adult, causing him to die of shock when the baby's body hits him in face. The newborn's actions will have played no role in bringing about the adult's death: this death is brought about simply by the newborn's existence in the world, as a passive lump of flesh, along with the gust of wind and other aspects of the surrounding causal architecture.

If such scenarios seem outlandish, that is simply an indication that the relevant risks are typically very small. Even when the relevant risks grow large, however, it is not obvious that we are obligated to internalize them. Consider

> *Waking From Coma:* Soon after he is born, little baby Gus slips into a coma. When he wakes up twenty years later, he is stunned to discover that his body is located atop of another person, Wanda, crushing her to death. Gus can roll his body off Wanda. But since they are both located atop a tall mountain, he will then fall to his death.

Gus' body is posing a lethal threat to Wanda, a threat which is in no way attributable to his agency. This threat is the materialization of a risk entailed by the very existence of Gus' body. And to me, at least, it seems initially plausible that Gus need not internalize the consequence of this risk; he need not roll off the mountain, killing himself, in order to prevent his body from killing Wanda.

If this judgment is correct - and I am not at all sure that it is - we may have an explanation of why certain astronomically small risk impositions should not generate obligations to internalize risk and strict defensive liability. For it might be suggested that an agent's risk impositions should have such moral significance only when they increase the amount of risk to which other people are subject, relative to the baseline of risk inherently entailed by the agent's very existence. Put simply, there is a certain quantum of risk - fixed by this baseline - that agents get to impose on others free of liability. And astronomically small risks of harm, this line of thought continues, fall beneath this baseline. Call this the *inherent risk* account.

The basic suggestion is interesting and it merits consideration. But it is not clear that it can be supplied with a principled rationale. If an agent need not internalize the risks of harm posed by his very existence, in cases like Waking From Coma, that is plausibly because the genesis of these harms is not at all connected to his responsible agency. The agent does not *act,* in giving rise to these harms, and so he does not infringe any duty in doing so. That is why, it seems plausible to suppose, the agent's duties to *prevent* or *undo* these harmful processes have at most limited stringency. So understood, the rationale for supposing that an agent need not internalize such risks does not support the conclusion that agents need not internalize astronomically small risks of harm *created by their actions* - even in cases where the relevant risks are no greater in magnitude than the risks imposed by passive existence. To support the point, consider

> *Waking From Coma*:* Soon after he is born, little baby Gus slips into a coma. When he wakes up twenty years later, he is stunned to discover that his body is located atop of another person, Ella, crushing her to death. Gus has three options. He can remain in place, in which case his body will crush Ella to death and he will suffer an excruciatingly painful death as well; he can roll to his left, in which case he will fall off a cliff and die painlessly; or he can roll to his right, in which case he will survive but trigger a bomb that will kill Ella.

It does not seem to me that Gus may trigger the bomb and kill Ella, although he will thereby impose on her a risk of death no greater than the risk of death imposed on her by his mere existence. While these risks are equal in magnitude, the former harm would be connected to Gus' agency in a way that the latter harm would not be. His duty against inflicting this lethal harm on Ella seems no less stringent than such duties typically are.

Even if the basic idea here is promising, moreover, I doubt it can underwrite a categorical exemption from strict defensive liability for astronomically small risk imposition. The relevant risks are incredibly small, but sometimes they are nevertheless larger than the risks entailed merely by existing in the world. To illustrate the point, compare two characters: a hermit whose sole waking activity is growing and

consuming the food necessary for his sustenance, within the four walls of his cave home,[209] and a hermit who enjoys taking strolls in the woods (among other healthy outdoor activities). The risk that a freak gust of wind will pick up the first hermit and toss him down a well, such that his body threatens to crush and kill another person located at the well's bottom, is astronomically small. The risk that a freak gust of wind will pick up the second hermit and toss *him* down a well, such that his body threatens to crush and kill another person located at the well's bottom, is *also* astronomically small. Nevertheless, the second risk is plausibly larger than the first. As surprising as it would be to hear of a freak gust of wind that picked up a man walking in the forest and hurled him down a well, it would be even more surprising (mind-boggling, even) to hear of a man subjected to a similar fate by a gust of wind that plucked him from within the walls of his cave.

Even if the inherent risk account can exempt the first hermit from strict defensive liability, therefore, it cannot exempt the second. Thus it cannot support a categorical exemption from strict defensive liability for astronomically small risk imposition.

V. Sharing the Expected Benefits and Burdens of Risky Action

As we have observed, an agent who dials his phone, as in Phone Operator, imposes extremely small risks of harm on many other people. Arguably, however, he *also* imposes extremely small expected *benefits* on each of these people. Just as there is some astronomically small chance that dialing a phone will trigger a causal chain that will kill one's neighbor, there is an astronomically small chance that doing so will trigger a causal chain that will *save* him from being killed; the same is true with respect to any other sort of harm, and with respect to any other person who stands to suffer it. And it is reasonably plausible, if not entirely obvious, that these risks are equivalent in magnitude, even if the magnitude of each risk is indeterminate.

These observations suggest an interesting approach to thinking about strict defensive liability for astronomically small risks of harm. The basic thought is that, when an agent imposes a risk of harm on another person, she may not be obligated to internalize that risk (and is not liable to defensive harm that will in effect force her to internalize it) if the risk is offset by an equally large expected *benefit* to that person.

[209] Inspired by the similar characters in H. Nye, "Chaos and Constraints," p. 15; and Ferzan, "Culpable Aggression," p. 682.

Call this basic idea the *expected benefit account.* The idea may require further refinement.[210] But it has some promise, for it does seem that, in various other ways, the presence of offsetting expected benefits can be morally significant.[211]

Our question, of course, is whether this factor is morally significant in a very particular way: whether it bears on an agent's obligations to internalize risk and his strict defensive liability. And we must keep in mind that, in cases like Phone Operator, there is another possibly relevant factor in place. The agent's action does not just impose astronomically small risks of harm and astronomically small expected benefits on other people*;* it imposes astronomically small risks of harm and astronomically small expected benefits on *himself*. For there is some astronomically small chance that an agent will (say) blow up his *own* house by dialing his phone, and some astronomically small chance that dialing his phone will prevent his own house from being blown up.

In order to test the expected benefit account, we need a case that will replicate this pattern of expected benefits and burdens. And the case should not deal in risks that are astronomically small, for we are trying to find an independently compelling explanation - one with more general application - of why the presence of astronomically small risks should be morally significant. Unfortunately, constructing an appropriate case proves rather complicated (so too, as an advance warning, will the rest of the discussion in this section). Here is such a case:

Gas Releaser: If Threatener releases a certain gas, it will have any (or all) of four possible effects:

- it may save Threatener from a potentially deadly heart attack which he is suffering (90% likelihood)

[210] Perhaps it should not apply, for example, if the relevant harm is death and the relevant benefit is a tasty meal, even if the expected value of risking each outcome is the same (because the chance of the former is so much smaller than the chance of the latter). Cf. the treatment of "pure benefits" in. S. Shiffrin, "Wrongful Life, Procreative Responsibility, and the Significance of Harm," *Legal Theory* 5 (1999) 2: 117 - 148. But the cases we are considering are not like that; in these cases, an astronomically small risk of harm is offset by an astronomically small chance of being saved from equivalent harm.

[211] To see this, suppose that an agent must, in order to achieve some very morally important end, press a blue button that will impose a 1% risk of death on Tomas or press a red button that will impose a 1% risk of death on Victoria. Pressing either button will also, with 1% likelihood, save Tomas from some different threat to his life. It seems plausible the agent should press the blue button: he should impose a risk of harm on the very person that his action might also save from harm.

- it may blow up Threatener's house, thereby killing him (10% likelihood)
- it may *prevent* Threatener's house from being blown up, thereby saving him (10% likelihood)
- it may blow up his neighbor Victim's house, thereby killing her (10% likelihood)
- it may *prevent* Victim's house from being blown up, thereby saving her (10% likelihood)

Threatener can internalize the 10% risk of death that he is imposing on Victim, by requesting that an appropriately situated impartial observer defensively monitor him with her sniper rifle, and kill him if and only if she discerns that by releasing the gas he is about to kill Victim. Threatener has no other way to internalize this risk.

It is permissible for Threatener to release the gas. But is he obligated to internalize the risk of death that he will impose on Victim by doing so? I am not sure, but I am inclined to say yes. After all, Threatener is gaining a much larger net expected benefit (a 90% chance of being saved from a deadly heart attack) from pressing the button than Victim is. Indeed, she is gaining no net expected benefit at all: the chance of benefit that Threatener is conferring on her is precisely offset by the risk of harm that he is imposing on her. Of course, if Threatener decides to internalize this risk, then Victim *will* be a net expected beneficiary of Threatener's decision to press the button (doing so will provide her with a 10% chance of avoiding death). Even then, however Threatener will be a much larger expected beneficiary than she will. It seems inequitable for Threatener to make himself a *very* large net expected beneficiary of his choice to put both of. their lives at risk, while leaving Victim with no expected benefit on balance at all, when he can instead make himself a somewhat smaller net expected beneficiary of action and allow Victim some net expected benefit on balance.[212]

[212] This conclusion might be supported by the following reasoning. Suppose that Victim calls Threatener before he presses the button, and demands that he internalize the risk he is imposing on her by doing so (thereby bringing about the more equitable division of expected net benefits from this risky action). It seems that this demand would have force, and that Victim would be permitted to give her consent to Threatener's action only on the condition that he internalize the risk in this way. To be sure, it may sometimes be permissible to risk harm to a person even when she explicitly withholds her consent from this risk imposition. But this does not strike me as such a case.

If this is correct, then the expected benefit account cannot establish a categorical exemption from strict defensive liability for threateners imposing only astronomically small risks of harm. For at least some such threateners will be analogous, in the relevant respect, to the threatener in Gas Releaser. Consider, for example,

> *Phone Operator With Vitally Urgent Call:* Threatener is suffering from a potentially deadly heart attack. He can call the local ambulance service with his phone. Doing so will have any (or all) of the following four effects:
>
> - it may save Threatener from the potentially deadly heart attack he is suffering (90% likelihood)
> - it may blow up Threatener's house, thereby killing him (astronomically small chance)
> - it may *prevent* Threatener's house from being blown up, thereby saving him (astronomically small chance)
> - it may blow up his neighbor Victim's house, thereby killing her (astronomically small chance)
> - it may *prevent* Victim's house from being blown up, thereby saving her (astronomically small chance)
>
> Threatener can internalize the astronomically small risk of death that he is imposing on Victim, by requesting that an appropriately situated impartial observer defensively monitor him with her sniper rifle, and kill him if and only if she discerns that by dialing his phone he will in fact be killing Victim. Threatener has no other way to internalize this risk.

If the expected benefit account does not relieve Threatener of his presumptive obligation to internalize the risk of harm he is imposing on Victim in Gas Releaser, then it does not relieve Threatener of his presumptive obligation to internalize the astronomically small risk of harm he is imposing on Victim in Phone Operator With Vitally Urgent Call. So too with the strict defensive liability of each threatener.

But suppose, for the sake of argument, that Threatener is *not* obligated to internalize the risk or liable to defensive killing in Gas Release, and is not (accordingly) obligated to internalize the risk or liable

to defensive killing in Phone Operator With Urgent Call*. Even so, I doubt the expected benefit account can support the categorical exemption from defensive liability that we are seeking. Consider

> *The Woman Behind the Curtain:* In order to treat a deadly heart attack from which he may be suffering, Billy must release a certain gas into the air. This gas will have any (or all) of the following potential effects on Billy and another person, Sally. Sally's mouth is muffled, and she is located behind a velvet curtain in front of Billy, so they cannot communicate. Here are the possibilities:
>
> - Billy is suffering a deadly heart attack, and releasing the gas will save his life (90% likelihood)
> - Billy is *not* suffering a deadly heart attack, and releasing the gas will kill him (10% likelihood)
> - Sally is suffering a deadly heart attack, and releasing the gas will save her life (10% likelihood)
> - Sally is *not* suffering a deadly heart attack, and releasing the gas will kill her (10% likelihood)
> - Sally is *not* suffering a deadly heart attack, and releasing the gas will do nothing to her (80% likelihood)
>
> As Billy knows, Sally is fully aware whether or not Billy is going to kill her by releasing the gas. However (as Billy also knows) she cannot communicate this to Billy, for she is muffled and hidden away behind the velvet curtain. In fact, it turns out, Billy *is* going to kill Sally by releasing the gas, and Sally knows this. Now an impartial observer can save Sally only by defensively killing Billy before he releases the gas.

It seems especially plausible to me, in this case, that Billy would be obligated to internalize the risk of death that he is imposing on Sally if he could - and that he is liable to defensive harm that will in effect force him to internalize it. It seems relevant that, while risking death to Sally also confers an expected benefit on her from *Billy's* perspective, it does not confer any expected benefit on her from *her* perspective. And Billy knows that, in the bad scenario where releasing the gas will in fact harm Sally, subjecting her to this risk imposition provides her with no offsetting expected benefit at all from her perspective.

This sort of fact - that a victim's harm is not offset by any expected benefit that she enjoys from *any perspective that the parties can share* - often seems morally relevant, as Alec Walen has recently observed in a

related context.[213] And if I am correct, in the present context, that Threatener would be obligated to internalize the risk of harm that he is imposing on Victim - and is liable to functionally equivalent defensive harm - then the same will be true in at least some cases involving astronomically small risk imposition. For example:

> *Phone Operator With Vitally Urgent Call (Victim Knowledge Variant)*: As in Phone Operator With Vitally Urgent Call. However, the villain who has programmed Threatener's phone, such that it will kill Victim when it is dialed, has also informed Victim about this nefarious scheme. She is now watching Threatener dial his phone (through a video feed that the villain has provided her), and knows that he is going to kill her when he finishes doing so. Observer knows all this, and can save victim only by shooting Threatener with a sniper rifle.

In the relevant respects - except for the astronomically small risks and expected benefits involved - the case is on a par with The Woman Behind the Curtain. So if my claims about The Woman Behind the Curtain are sound, then we cannot invoke expected benefit reasoning to explain why astronomically small risk impositions need not be internalized and do not generate strict defensive liability in cases such as this.

CONCLUSION

The cases and considerations discussed in the last section hover on the edge of prohibitive and faintly ludicrous complexity. My intuitive judgments and arguments about these cases are not obvious or irresistible, moreover, and further reflection might exert pressure in the opposite direction. For those especially keen to exempt astronomically small risk imposition from strict defensive liability, therefore, this might be a promising place to press further. But pressing further is likely, of course, to involve even more prohibitive complexity.

This somewhat dispiriting observation leads naturally to a more general methodological point. I have argued that we require some principled basis to exempt astronomically small risk impositions from generating risk internalization obligations and strict defensive liability. But we have seen that the available

[213] A. Walen, "Risks and Weak Aggregation: Why Different Models Suit Different Types of Cases," *Ethics* 131 (2020) 1: 62 - 86.

rationales are unlikely to support any such categorical exemption. And even if a categorical exemption can be secured, it is likely to rest on dauntingly difficult and intricate philosophical argumentation - argumentation of a kind that may be substantially different from anything that philosophers working on the morality of the defensive harm have until now entertained. Even if the relevant moral considerations can underwrite the desired categorical exemption, therefore, these considerations may prove sharply discontinuous with the basic phenomenology of our intuitive convictions about this moral domain. It is difficult to escape the suspicion that these intuitive phenomena are best characterized in a very basic way, one that seems indefensible on even a modicum of reflective scrutiny: astronomically small risk impositions *are simply too small,* or too freakishly unusual in character, to possess the moral significance that risk impositions otherwise possess.

It is unimaginable, however, that any such proposition could carve at the moral joints, on anything like the robustly realist picture of morality which philosophers working on defensive liability tend to explicitly embrace or implicitly presuppose.[214] If by contrast the fundamental moral norms are best characterized in an anti-realist or constructivist fashion - if they are ultimately grounded in our messy and arbitrary patterns of moral thinking and feeling,[215] or if they are at least partly constructed from the contingent circumstances of human history and life[216] - then it would make greater sense that the morality of defense and risk internalization should make such an arbitrary carve-out for astronomically small risk imposition. There are exceptional circumstances - such as the pandemic that is killing thousands of people as I complete this article - in which such risks assume widespread psychological and deliberative salience. But in the ordinary course of human affairs, such risks are confined to the dark peripheries of our consciousness. So there is typically little occasion for our cognizance of such risks to shape our moral rights and duties, in the way that an anti-realist or constructivist picture of morality would envisage.

---

[214] See e.g. J. McMahan, "Moral Intuition," in *The Blackwell Guide to Ethical Theory* eds. H. LaFollette and I. Persson (West Sussex: Blackwell Publishing, 2013): 103 – 123.
[215] See e.g. S. Street, "Objectivity and Truth: You'd Better Rethink It," in *Oxford Studies in Metaethics* vol. 11, ed. R. Shafer-Landau (Oxford: Oxford University Press, 2016)*:* 293 - 334.
[216] See. e.g. K. Setiya, "Humanism," *Journal of the American Philosophical Association* 4 (2018) 4: 452 - 470.

Thinking about astronomically small risk imposition, and laboring in the weeds of deontological ethics more generally, has certainly made me more sympathetic to anti-realist and constructivist pictures of the moral domain than I would otherwise be. But on balance I remain skeptical of such pictures and believe there are serious philosophical obstacles to their acceptance. And for the reasons I have given, I have become increasingly confident that any realist picture will simply need to accept that certain astronomically small risk impositions do indeed generate strict liability to defensive harm. On balance, therefore, I suspect we should simply accept this conclusion, even if our initial intuitive sensibilities rebel against it.

By way of conclusion, let me offer three novel and mutually reinforcing arguments for this conclusion, which are mercifully brief. Each involves a variant of Phone Operator With Urgent Call. First, consider

> *Double Prevention:* Threatener is dying of a heart attack and needs treatment to survive. He is about to dial his cell phone, in order to call an ambulance to take him to the hospital. But a villain has programmed Threatener's cell phone such that it will blow up Victim's house and kill her once it is dialed. Observer can save Victim only by using a sniper rifle to shoot the cellphone out of Threatener's hands.

It seems to me intuitively that Observer may shoot the cell phone out of Threatener's hands. But the foreseeable consequence of this intervention is that Threatener will die and Victim will survive. Observer would produce the same consequence - though obviously in a different way - by just defensively shooting and killing Threatener, in order to save Victim, as in Phone Operator With Urgent Call. There is at least some tension between supposing that the former intervention is permissible while the latter intervention is impermissible.

Consider, next,

> *Communication of Certain Knowledge:* Threatener is dying of a heart attack and urgently needs medical treatment to survive. He is about to dial his cell phone in order to call an ambulance to take him to the hospital. But a villain has programmed Threatener's cell phone such that it will blow up Victim's house and kill her once it is dialed, as Observer knows with (near) certainty. Observer can save

Victim by calling Threatener on his landline and informing him that using his cell phone will kill Victim. The foreseeable consequence of this intervention will be that Threatener will die. For either he will then comply with his obligation to refrain from using his cellphone, and succumb to his deadly heart attack, or he will culpably refrain from complying with this obligation, in which case Observer can (and will) defensively kill him with a sniper rifle.

It seems to me intuitively that Observer may call Threatener and, if Threatener still chooses (now culpably) to use his cellphone, defensively shoot and kill Threatener with a sniper rifle. But the foreseeable consequence of this initial intervention is that Threatener will die - no less than if Observer, finding himself unable to communicate with Threatener, just defensively shoots and kills him. So it is a bit strange to think that the former course of action is permissible while the latter is not.

Consider, finally,

Communication of Uncertain Possibility: Threatener is dying of a heart attack and urgently needs medical treatment to survive. He is about to dial his cell phone in order to call an ambulance to take him to the hospital. As Observer knows, however, there is a 1% chance that a villain has programmed Threatener's cell phone, such that it will blow up Victim's house and kill her once it is dialed. Observer can call Threatener on his landline and inform him of this fact. Threatener will still be permitted to use his cell phone, since the risk he will thereby impose on Victim will be reasonably small. But he will then be obligated to internalize this risk, by asking Observer to defensively monitor him and kill him if necessary to prevent him from killing Victim; and if Threatener does not honor this obligation, he will be liable to this defensive measure (as a culpable threatener) at Observer's hands.

It seems to me intuitively that Observer may call Threatener's landline and inform of him of the possibility that he will be killing Victim by dialing his cellphone. But the foreseeable consequence of this intervention is that Observer will then permissibly subject Threatener to defensive monitoring and killing (if necessary). It is at least somewhat strange to suppose that the permissibility of forcing Threatener to internalize the potentially harmful effects of his action, in this way, should turn on whether Observer can successfully

communicate with Threatener - which might itself turn on such factors as whether the landline is rickety, or whether Threatener and Observer speak in different idiolects.

In each case it is strange, therefore, to suppose that Observer may permissibly engage in *one* course of action that will foreseeably bring about Threatener's death (shooting the cell phone out of his hand; informing him about the certain effect of his action on Victim and then defensively killing him if necessary; informing him about the possible effect of his action on Victim and then defensively monitoring and killing him if necessary) but that Observer cannot engage in *another* course of action that will foreseeably bring about the very same consequence (i.e. defensively killing him straightaway, if necessary.) In one respect, at least, it is plausible to maintain on reflection that *all* such courses of action are permissible or *none* of them are. If one of these uniform treatments must be chosen, the former strikes me as preferable. And to adopt this uniform treatment is to maintain that Threatener is liable to defensive killing in all of these cases, although the risk of harm he is imposing on Victim is astronomically small.

The force of the point should not be overstated. It is a general feature of deontological ethics, which arises ubiquitously and in many manifestations, that it may be permissible to harm a person in one way but not in another. Still, the point has some force. It lends some modest support to a conclusion that may be inescapable on other grounds: imposing astronomically small risks of harm on other people, in even the most quotidian ways, can render us liable to severe defensive harm.