2005

Responsibility for Unintended Consequences

Claire Oakes Finkelstein
University of Pennsylvania, cfinkels@law.upenn.edu

Responsibility for Unintended Consequences

Claire Finkelstein*

The appropriateness of imposing criminal liability for negligent conduct has been the subject of debate among criminal law scholars for many years. Ever since H.L.A. Hart's defense of criminal negligence, the prevailing view has favored its use. In this essay, I nevertheless argue against criminal negligence, on the ground that criminal liability should only be imposed where the defendant was aware he was engaging in the prohibited conduct, or where he was aware of risking such conduct or result. My argument relies on the claim that criminal liability should resemble judgments of responsibility in ordinary morality as closely as possible. I argue that responsibility judgments in ordinary morality are based on the agent's having acted intentionally, and that an agent does intentionally what he chooses to do. Because agents choose to bring about those effects of their actions they foresee as reasonably likely to follow from what they do, they are responsible for such effects. They are not responsible for effects they do not foresee, or for effects they deem highly unlikely, and they ought not to be held criminally liable for them either.

I.

The purpose of this paper is to revisit an old question, namely whether there should be crimes of *negligence*. The anti-negligence position was most famously articulated by J.W.C. Turner in 1936, in an essay arguing that foresight of criminal harm provides a necessary condition for criminal liability.[1] Turner's primary argument for this claim was that a system of criminal liability that dispenses with foresight of harm is tantamount to a system of strict liability. This position was commonly accepted among criminal law theorists for many years.[2]

Turner's position, however, was forcefully attacked by H.L.A. Hart, who argued that the idea that negligence is a form of strict liability is based on "a

---

* Professor of Law and Philosophy, University of Pennsylvania. I wish to thank Antony Duff, Leo Katz, Andrew Simester, and members of the audience at the 2003 IVR workshop where this paper was originally presented, as well as members of the UCLA Legal Theory Workshop. I owe a particular debt of gratitude to Kurt Baier, David Gauthier, and Michael Thompson for their patient supervision of the original doctoral dissertation, entitled *Ethics and the Intentional*, from which Parts II through V of this paper is drawn. I also benefited from conversations with George Fletcher on the question of criminal negligence.

[1] J.W.C. Turner, *The Mental Element in Crimes at Common Law*, 6 CAMBRIDGE L.J. 32 (1936); *see also* J.W.C. TURNER, THE MODERN APPROACH TO CRIMINAL LAW (1945).

[2] *See* Jerome Hall, *Negligent Behavior Should Be Excluded From Penal Liability*, 63 COLUM. L. REV. 632 (1963).

mistaken conception both of the way in which mental or 'subjective' elements are involved in human action, and of the reason why we attach the great importance which we do to the principle that liability to criminal punishment should be conditional on the presence of a mental element."[3] Turner's mistake, Hart claimed, was to see negligence as a state of mind. While he agreed that the mind of the man who acts inadvertently is a blank with respect to something he did negligently, Hart argued that we need not locate the notion of *mens rea* in a defendant's subjective state of mind. Instead he writes, "we can perfectly well both deny that a man may be criminally responsible for 'mere inadvertence' and also deny that he is only responsible if 'he has an idea in his mind of harm to someone.'"[4] Negligence lies not in the state of a man's mind, but in his failure to live up to an objective standard of conduct.

Imagine a workman mending a roof in a busy town, throwing bricks into the street below him without looking to see if anyone is passing by. According to Hart, Turner would require us to choose between two possibilities: Either the workman has the conscious idea that he might harm someone, in which case he is to blame for any injuries he causes, or he has no such idea, in which case any harm he inflicts is inadvertent. But, argues Hart, the workman "failed to comply with a standard of conduct with which any ordinary reasonable man *could* and *would* have complied: a standard requiring him to take precautions against harm."[5] This entitles us to hold him responsible, despite the fact that he lacked awareness of what he was doing. Hart's central argument is thus that we can account for the "subjective" element of fault in normative terms, by tying the standard of liability to an individual's particular capacities. The workman acts negligently, not merely inadvertently, because he has a duty to take care to avoid injury, and he has the capacity to take such care. As long as the standards to which we hold one another are adjusted to account for the capacities of the defendant, there is no reason to think we have adopted an "objective," or strict standard of liability.

Criminal law theorists have largely sided with Hart's view of the matter, with the result that Turner's position is no longer seriously defended. Most criminal commentators now seem to accept liability for negligence in at least some form.[6] In my view, however, Turner had the better position, even if he lacked compelling arguments for it. There are two reasons for this. First, Hart incorrectly takes Turner's argument to be a blanket attack on objective systems of liability. Hart then counters what he takes to be Turner's point by turning to a standard that takes into account the defendant's own capacities. And Hart is right that a negligence standard tailored to the defendant's capacities would certainly be non-objective in

---

[3] H.L.A. HART, PUNISHMENT AND RESPONSIBILITY: ESSAYS IN THE PHILOSOPHY OF LAW 139 (1968).

[4] *Id*. at 147.

[5] *Id.* at 147–48.

[6] *See* George Fletcher, *The Theory of Criminal Negligence: A Comparative Analysis*, 119 U. PA. L. REV. 401, 415 (1971).

the sense of being *individualized*. But *individualized* is not the same as *subjective*. And Turner's argument is not just an attack on objective approaches to liability. It is rather an insistence that a person should not be held criminally liable in the absence of some subjective awareness on his part of what he was doing, under its prohibited description. To answer this challenge by saying that negligent liability is individualized is non-responsive.

Second, and more importantly, Turner's claim that negligent liability is a form of strict liability seems to me to be absolutely correct. It is, moreover, correct for precisely the reason Turner gives, namely that liability in the absence of that subjective element is objectionable because it is liability in the absence of ordinary responsibility. Turner's problem was that he was unable to offer any account of responsibility, and hence he lacked any way of accounting for the relevance of subjectivity to that notion. It is the purpose of the present essay to remedy this deficiency by providing such an account.

In what follows, I approach the question by placing negligent criminal liability in the context of a more general theory of responsibility for action. That is, I assume that the criminal law is simply another form of ordinary, non-criminal responsibility. (I attempt some defense of this claim at the end of the essay, but for the moment will simply assume it.) I then approach the question of negligent responsibility in a roundabout way: I examine another type of responsibility for "unintended" effects, namely cases in which an agent is aware of, but does not intend, the violation of a prohibitory norm. Extrapolating from these cases, I argue that our ordinary responsibility practices are predicated on the notion of choice. As such, they extend only to things agents do with awareness of what they are doing or risking. I conclude that negligence is incompatible with traditional principles of criminal responsibility.

II.

Placing criminal liability in the context of judgments of responsibility in ordinary morality does not make our task easier. For ordinary morality both seems to encompass, and to reject, liability for negligent conduct. We sometimes find ourselves irate with a person for forgetting something important, at the same time that we accept "I simply forgot" as a plea in exoneration. While criminal law theorists have sometimes thought ordinary morality firmly on the side of responsibility for inadvertence, ordinary moral practice is not as clear a guide as might be supposed. For this reason, I shall approach the problem of negligent responsibility by the back door. Instead of considering ordinary moral practices relating to forgetting, inadvertence,[7] and failure to take precautions, I shall begin by considering our moral practices relating to clear cases of responsibility. I shall

---

[7] *See id.* at 415 ("In daily conduct, we confidently blame others who fail to advert to significant risks.").

then attempt to extrapolate a conception of responsibility from these more ordinary cases, and see what it in turn implies for cases of negligence.

The dominant treatment of responsibility in the philosophical literature sees that notion as a normative one—a judgment made of a human agent for something he did, rendered on moral grounds. Moreover, normally the thing the agent did was bad or reprehensible. Philosophers treat an agent as "morally responsible" for what he is to *blame* for having done. They do not tend to speak of "responsibility" for *praiseworthy* acts.

This way of proceeding may seem peculiar, but many authors on this subject at least tacitly assume it, and some even defend it explicitly. R.J. Wallace, for example, writes that

> the question of what it is to be a morally responsible agent should be given what I call a normative interpretation. If we wish to make sense of the idea that there are facts about what it is to be a responsible agent, it is best not to picture such facts as conceptually prior to and independent of our practice of holding people responsible.[8]

Hart had previously defended this position as well. In *The Ascription of Responsibility and Rights*, he wrote: "[S]entences of the form 'He did it' have been traditionally regarded as primarily descriptive whereas their principal function is what I venture to call *ascriptive*, being quite literally to ascribe responsibility for actions."[9] Responsibility ascriptions are made, he claimed, in accordance with a set of antecedent moral or legal norms. Hart's account in that paper is both *prescriptive* and *conventional*: Responsibility is something that must be assigned, the way people are assigned numbers waiting on line in a bakery, and the form in which responsibility is assigned is a function of social purposes. As Hart wrote, "assigning responsibility in the way we do assign it tends to check crime and encourage virtue," and Hart thought this provides both the explanation and the justification for our assigning responsibility where and how we do.[10]

While Hart later distanced himself from his early work on responsibility,[11] the ascriptivism of his account remained in other guise, and with it, the problems with that approach. The central problem is that the ascriptive approach leaves us empty-handed when it comes to describing an agent's relation to ordinary, non-morally charged actions. Smith's playing the piano, on this account, is not something we can lay at Smith's door, assuming he did not play particularly well or badly, since there would be no moral or conventional purpose to making such an

---

[8] R.J. WALLACE, RESPONSIBILITY AND THE MORAL SENTIMENTS 1 (1994).

[9] H.L.A. Hart, *The Ascription of Responsibility and Rights*, *in* LOGIC AND LANGUAGE 151 (Antony Flew ed., 1965).

[10] *Id.* at 173. This aspect of Hart's view remained central to his writing in later years, and to his famous claim that what he called the General Justifying Aim of Punishment is a utilitarian one.

[11] *See* HART, *supra* note 3, at 145.

assignment. But we would say that Smith is responsible for injuring the child, if he hit him, or for breaking the window, if he threw a baseball through it, since judging him thus fits with some normative purpose we have in mind. It is odd, however, to think of playing the piano as fundamentally different from hitting the child or breaking the window. We most naturally think of agents as equally responsible in all three cases.

What we require from a notion of responsibility is a way of capturing an agent's relation to ordinary things he does, without yet considering whether those actions are morally significant.[12] Thus Smith is the agent of, or is responsible for, playing the piano, tying his shoelaces, combing his hair, and sipping his coffee. Moreover, he is responsible for these things in a way that he is not responsible for tripping over the rug, spilling his coffee, and kicking his leg when the doctor tests his reflexes. What we want, in other words, is a way of capturing an agent's relation to things he does *qua rational agent*, since these are things for which it makes sense to raise questions of praise or blame. We require, in short, a descriptive conception of responsibility.

In criminal law theory, the persistent application of a normative conception of responsibility has had certain infelicitous results. First, the normative approach makes it difficult to make sense of certain justifications and excuses. What should we say, for example, about the person who robs the bank under duress, if we think the pressure he was under provides him with a good excuse? Most naturally, we would say he is responsible for robbing the bank, since he did so intentionally, but that we do not blame him for it because he did it for a good reason. That is, we distinguish his *prima facie* responsibility for robbing the bank from his *ultimate* blameworthiness for it, and find a basis for withholding criminal punishment based on the latter. On a normative conception of responsibility, however, we must say he is not *responsible* for robbing the bank. Not surprisingly, this is precisely how Hart treats such cases: He says that the fact that someone did something "accidentally," "inadvertently," "by mistake," or "while insane" has the same effect as his acting in self-defense or under duress, in that both sorts of defense defeat prima facie responsibility.

The question regarding negligence arises at the level of the prima facie case: Is a person who brings about harm inadvertently responsible for it, in the descriptive sense? Hart says that a person can be responsible for things he does negligently because we can blame him for failing to live up to a normative standard of which he was capable. But this approach confuses responsibility with blameworthiness. We cannot ask about a person's normative failings until we examine whether he is responsible in a descriptive sense. Our question is thus whether people are responsible for what they do inadvertently, such that they might sometimes be blamed for failing to live up to a certain normative standard.

---

[12] I mean here to be focusing on responsibility for "things done," as opposed to what we might call "capacity responsibility," namely the set of capacities that characterize normal adults.

This will require us to investigate more thoroughly the nature of the descriptive notion of responsibility for which I have been arguing. We turn to this below.

A second drawback of the normative approach to responsibility is that it will require divergent accounts of responsibility for moral and legal judgments. Since moral and legal criteria for normative judgments diverge, the relevant accounts of agency will also diverge if that notion is already infused with judgments of blameworthiness. But it seems reasonable to suppose that the conditions of responsibility themselves should be invariant as between moral and legal responsibility, and that what differs between the two is just the particular normative system one brings to bear on the responsible agent's conduct. Moral and legal responsibility should have a *common denominator*, and it is this that the normative approach to responsibility cannot capture. The task of the next section will be to articulate an account of a descriptive notion of responsibility that might serve as this common denominator.

III.

What is the feature that best characterizes things human agents do in their capacity as rational animals? The usual philosophical answer is that human beings act as rational animals insofar as they act according to an end. Human reason is thus fundamentally teleological, and human behavior is characteristically rational insofar as it is motivated in this way. The notion of responsibility we have been discussing is both characteristically human and intrinsically related to the idea of action. A natural thought to have about the notion of responsibility we have identified, then, is that it can be accounted for in teleological terms.

We may find confirmation for this approach in the traditional philosophical analysis of the notion of "intentional action." According to the Standard Account, a person acts intentionally just in case he acts *for a reason*, where *acting for a reason* implies that he acts for the sake of something he wants or is trying to get. The Standard Account can be summarized with the following two theses:

> **Thesis (1)**: Someone does something intentionally if (and only if) he does it for a reason.
>
> **Thesis (2)**: Someone does something for a reason if (and only if) he does it for the sake of something he wanted or was trying to get.

We might then attempt to combine the Standard Account with a thesis about responsibility:

> **Thesis (3)**: Someone is responsible for something he did if (and only if) he did it intentionally.[13]

---

[13] J.L. Mackie calls a closely related principle the "straight rule" of responsibility. J.L.

In combination with the Standard Account of intentional action, Thesis (3) would define the ambit of responsibility as those things an agent does for the sake of something he wants.

If all three of the above theses were correct, we would have to conclude that agents are not responsible for things they do unintentionally. But it will be immediately apparent that we cannot reason in this way, since there are at least some cases of things agents do unintentionally for which we *cannot* deny responsibility. Considering such cases will make clear that at least one of Theses (1) through (3) must be false.

Consider this case. A seller of goods is about to ship some heavily insured cargo aboard a passenger plane, when the buyer cancels his order for the goods. Faced with the prospect of imminent financial ruin, the seller formulates a plan to destroy the plane: plant a bomb to explode mid-flight and collect the insurance on the goods. He regrets that the passengers will almost certainly die in the process, but he is not dissuaded. He executes his plan, and as expected the cargo is destroyed, and no one survives. Call this case "Insurance Bomber."

Unlike where examples of negligence are concerned, we cannot simply abandon the claim that the insurance bomber is responsible for the deaths for the passengers, given that he is fully aware his actions will result in their deaths. Indeed, Insurance Bomber is a paradigmatic case of responsibility for evil, since truly reprehensible conduct often stems more from indifference and selfishness than from directed malevolence. But it should be clear that we cannot say the insurance bomber is responsible for the deaths of the passengers at the same time that we maintain Theses (1) through (3). For Thesis (3) implies that if the bomber is responsible for killing the passengers, he must have killed them intentionally. But according to the Standard Account (Theses (1) and (2) together), he did not kill them intentionally because he did not kill them *for a reason*. We must therefore choose between rejecting the Standard Account and rejecting the thesis that connects responsibility with what an agent does intentionally (Thesis 3).

The obvious solution might seem to be to reject Thesis (3). And the instances in ordinary morality in which we *do* appear to blame agents for causing harm unintentionally weigh on the side of this solution. Cases like Insurance Bomber, we might be tempted to argue, show conclusively that we *do* hold agents responsible for things they do unintentionally. And while this does not mean that we hold them responsible for all unintentional harm they cause, there is no reason in principle to reject responsibility for unintentional harm in cases in which ordinary morality seems to support such ascriptions of responsibility. There is thus no reason in principle to reject the blameworthiness of negligent conduct in

---

MACKIE, ETHICS: INVENTING RIGHT AND WRONG 208 (1977). Mackie identifies the straight rule as the rule that "an agent is responsible for all and only his intentional actions." Thesis (3) talks about something someone did, rather than about actions, in order to include intentional omissions within its scope.

ordinary morality, and so no reason to reject the use of negligent liability in the criminal law.

Proponents of the Standard Account might argue in this way. Indeed, Jennifer Hornsby has argued vigorously that one should not expect the theory of action to "deliver the goods" for the criminal lawyer. On the one hand, the notion of intentional action is entirely a psychological one. She writes: "whether someone did something intentionally is, in a certain sense, a question about her . . . . [I]t is to her states of mind that we need to advert in order to settle the question."[14] But the question of responsibility is a normative one. There is thus no reason to suppose that the former concept would provide a basis for making ascriptions of the latter sort. Thesis (3) is confused, she would say, in seeking to connect the concept of intentional action with judgments of responsibility. Intentional action, like intention, tracks reasons for acting. But ascriptions of responsibility are entirely a matter of the norms to which we subscribe. Hornsby's view of intentional action would fit well with Hart's ascriptive account of responsibility.

I nevertheless believe this quick and straightforward solution to our problem mistaken, for at least two reasons. First is a point in ordinary language concerning the adverb "intentionally." It seems perfectly appropriate to say that the person who knocks over the vase through inadvertence does not knock it over intentionally. This is even so when the consequences are grave, such as the tragic case of the person who runs his own child over when backing down a driveway. If the agent was not aware of any risk that his child was behind him, then even if he *should* have been aware of a risk, we do not say he ran over his child "intentionally." But it would be exceedingly odd to say that the bomber did not kill the passengers intentionally. It is much less odd to say that he did not *intend* to kill them. And this indicates a curious feature about the family of concepts surrounding intention and intentional action: What a person does, *intending* to do it, may identify a narrower class than what a person does *intentionally*. The bomber, for example, did not intend to kill the passengers. But it seems quite natural to say he killed them intentionally.

Now the fact that more things are done intentionally (under a given description) than are done with an intention is not, by itself, an argument for retaining Thesis (3). But it does make it *possible* to retain Thesis (3), even if the reach of responsibility is broader than the class of things people do with an intention. And if the observation from ordinary language we made is correct, then there is reason to draw out the concept of what is done intentionally in precisely the way that is required: We can make the concept of intentional action as broad as the concept of responsibility. If we have a way of making this broader notion of intentional action philosophically respectable, we would have found a way of making Thesis (3) philosophically respectable as well.

---

[14] Jennifer Hornsby, *On What's Intentionally Done*, *in* ACTION AND VALUE IN CRIMINAL LAW 55, 66 (Stephen Shute et al. eds., 1993).

The second problem with the above argument for the elimination of Thesis (3) is that there is some independent reason to think Thesis (3) correct. In particular, cases like Insurance Bomber support the idea of a linkage between intentional action and responsibility. There is a felt connection between responsibility and what is done intentionally that is lacking in the relation between responsibility and what is done with an intention. Imagine how outrageous it would be for the bomber to say, "Yes, I admit I killed the passengers intentionally, since I knew that they would die. But I am not responsible for killing them, since I didn't intend to do so." Rather, the fact that he killed them intentionally seems to carry with it the idea that he is responsible for killing them. And conversely, it would be only a little less odd for someone to deny that he did something intentionally, but not regard that as mitigating his responsibility. The person who says, "I didn't do it intentionally" seems to be making a plea for exoneration. He seems to be offering his hearers a reason why they should not blame him for something he did.

To some, the idea of a connection between judgments of intentional agency and judgments of responsibility seems quite obvious. Indeed, some philosophers have claimed that the concepts bear an analytic relationship to one another. R. A. Duff, for example, writes that "[a]scriptions of intentional agency are, as a matter of meaning, ascriptions of responsibility."[15] If this intuitive connection between responsibility and intentional action is correct, we have particular incentive to try to cash out the latter notion in a way that will provide an appropriate foundation for judgments of responsibility. As we shall see, this would supply quite a different solution to our problem than rejecting Thesis (3), which seemed at first most appealing. Instead of rejecting Thesis (3), we would be required to reject either Thesis (1) or Thesis (2). Let us consider these possibilities in turn.

The solution that rejects Thesis (1) would preserve the view that acting for a reason is acting for the sake of an end. It would also retain the connection between responsibility and what an agent does intentionally. But it would reject the connection between what an agent does intentionally and what he does for a reason, and along with it, the standard association between what an agent does intentionally and what he intended to do. Acting with an intention, on this view, is associated with acting for a reason, while what an agent does intentionally would be linked to the broader category of responsible agency. Reason thus sides with the narrower concept of intention, rather than the broader concept of responsibility. According to this solution, then, the insurance bomber killed the passengers intentionally, and he is responsible for having done so, but he did not intend to kill them, since he did not kill them for a reason.

Michael Bratman favors this solution. He rejects what he calls the "Simple View," according to which it follows from the fact that someone did something intentionally that he intended to do it. Thus consider the person who knowingly scratches a car next to him when pulling into a tight parking space. The Simple View would say that if he intentionally scratched the car, he must have intended to

---

[15] R.A. DUFF, INTENTION, AGENCY & CRIMINAL LIABILITY 77 (1990).

scratch it. But since the driver did not intend to scratch the car, given that he did not *try* to scratch it, we cannot say he scratched it intentionally.[16] But Bratman thinks it is possible to do something intentionally without having intended to do it. The driver intentionally scratched the car, despite the fact that he never intended to scratch it. Bratman does not claim that *all* foreseen side effects of one's intentional actions are intentional. Only those effects that lie within the "motivational potential" of one's action should be thought of as done intentionally,[17] meaning that a person must have consciously adverted to and actually deliberated on an effect for it to count as something done intentionally. There is no reason to think that I intentionally wear down the soles of my shoes when I run a marathon in the normal course of events, even if I am aware of wearing them down. But if there is some reason why I particularly attend to that effect of my action, such as that my shoes are a family heirloom, then it makes sense to say I wear them down intentionally, despite the fact that I do not intend to wear them down. Such cases falsify the Simple View.

Someone who rejects the Simple View must reject the Standard Account as well: Theses (1) and (2) in combination commit one to the position that doing something intentionally entails that one intended to do it, assuming we treat acting with an intention as a species of acting for the sake of an end. So Bratman must reject either Thesis (1) or Thesis (2). Which of the two theses does he in fact reject? Bratman accepts the standard, instrumental approach to practical reason, according to which an agent acts rationally insofar as he acts in pursuit of something he wants. His argument for rejecting the Simple View is limited to the notion of intentional action, and he otherwise aligns *acting with an intention* with what an agent does *for a reason*. His central claim is that the usual account of practical reason fails to accord sufficient weight to planning activity. Plans give us reasons to do things, in a way that allows rational agents to avoid having to go back each time to their background reasons in order to decide what to do. But plans, like intentions, are teleological: an agent who acts on a plan acts for the sake of an end. As Bratman explains, plans are "intentions writ large."[18] By expanding rational agency to include planning, Bratman is not fundamentally challenging the teleological structure of instrumental rationality. His rejection of the Simple View is thus an implicit rejection of Thesis (1).

For the driver pulling into the tight parking spot, Bratman says, scratching the adjacent car is part of his overall plan for parking his car.[19] Its being part of his plan makes it intentional, but it does not make it something done for a reason, since scratching the car is not a *means* to parking the car. On both the standard conception and the planning conception of agency, the driver has no *independent*

---

[16] MICHAEL E. BRATMAN, INTENTIONS, PLANS AND PRACTICAL REASONS, ch. 8 (1987).

[17] *Id.* at 124–26. For a variation on this theme, see DUFF, *supra* note 15, ch. 4.

[18] BRATMAN, *supra* note 16, at 29.

[19] *Id.* at 29.

reason to scratch the car—independent of his plan for parking his car. And if he has no *independent* reason for scratching the car, he has no reason for scratching it, Bratman says.

But someone might argue for a different understanding of such cases. Why does the fact that the driver has no *independent* reason to scratch the adjacent car mean he has no reason? Why doesn't the fact that he has a reason for adopting the plan of which scratching the car is a part give him a reason to scratch it? True, if he pulled into the space without scratching it, he would not have a reason to back up and try again, this time cutting it closer so as to ensure a scratch. But it is not clear why *independence* is a necessary condition for something to count as a reason. Another feature of reasons is that they provide explanations for people's actions. Here, the fact that the driver wanted to pull into the parking spot does explain *why* he scratched the car. I shall explore this thought in greater detail in the next Part. I mention it now only to suggest that if plausible, it would provide us with a way of making sense of Bratman's claim that the driver scratches the car intentionally—namely that he does so *for a reason*.

The problem with accounts that reject Thesis (1) is that it is not of any obvious benefit to be able to say that an agent did something intentionally if that notion is not one we can relate to the rational-explanatory principle on which he acts. The intuition that agents are responsible for what they do intentionally presumably stems from an intuitive link between an agent's doing something intentionally and his reasons for acting. Judgments of responsibility bear a special relation to rational agency to the extent they bear on an agent's reasons for acting. To misappropriate a thought from Elizabeth Anscombe, the realm of responsibility is the realm of things to which a certain sense of the *Why?* question has application, where that question calls for the agent's reason for doing what he did.[20] But if we reject the connection between intentional action and reasons for acting, we cannot say that a person who acts intentionally must be able to explain what he did in terms of his reasons for acting.

Let us call an account of responsibility that connects what an agent is responsible for with what he does for a reason an "internalist" account. And let us call any account that denies this connection "externalist." The view that consists in rejecting Thesis (1) and retaining Theses (2) and (3) is externalist, since the category of what an agent does intentionally is broader than the category of what is done for a reason. What I have in effect suggested is that we may have grounds for preferring an internalist to an externalist account, insofar as judgments of responsibility are most intelligible to us if they connect with reasons for acting. Let us consider what an internalist account, namely an account that rejects Thesis (2) and retains Theses (1) and (3), might look like.

V.

[20] ELIZABETH ANSCOMBE, INTENTION § 5 (1957).

Insurance Bomber demonstrated a tension between our intuitions about responsibility and our intuitions about intentional action. We have considered two possible solutions thus far. The first is the Standard Account of intentional action, according to which the notion of responsibility should have nothing to do with our underlying account of intentional action. The second solution would distinguish what a person does with an intention from what he does intentionally, as a way of allowing judgments of responsibility to connect with the latter notion. But it would restrict the category of things people do for reasons to that which they intended to do, and would articulate the category of intentional action in a way that is distinct from reasons for acting.

We will now consider a third solution. This solution preserves the connections between what an agent does intentionally and responsibility, as well as between what an agent does intentionally and what he does for a reason. Its account of responsibility is "internalist," insofar as it ties judgments of responsibility to reasons for acting. But it abandons the claim that acting for a reason is limited to acting for the sake of an end. Like the second solution we considered, this solution uncouples doing something intentionally from doing something with an intention: What is done with an intention is still understood in terms of the "for the sake of" relation, while what is done intentionally is accounted for in terms of some broader view of acting for a reason.

The difficulty for this account lies in offering an acceptable alternative to the "for the sake of" requirement Thesis (2) imposes. Reasons for acting have long been thought of in teleological terms—the reason *for* which something is done is the reason for the sake of which it is done. If we wish to say the insurance bomber kills the passengers for a reason, we will have to offer an account of what it is to do something for a reason which does not require reason *for the sake of which*, since the insurance bomber does not kill the passengers for the sake of anything.

The insurance bomber's reason for blowing up the plane is *to get the insurance money on the cargo*. If asked *why* he blew up the plane, his answer, if truthful, would mention this reason. The claim of those who adhere to the Standard Account is that unlike blowing up the plane, the insurance bomber has no reason for killing the passengers. As Anscombe suggests, the *Why?* question in this case is "refused application."[21] But as we already briefly saw in discussing Michael Bratman's view, perhaps this is incorrect. Suppose we were to ask the insurance bomber *why* he killed the passengers. Although he cannot answer with "in order to get the insurance money," he might respond by saying "Well, I wanted the insurance money, and to get it I had to blow up the plane and kill the passengers." Compare this *Why?* question with a question about something else the insurance bomber might have done, namely *run the insurance company out of business by creating such a large claim*. Assuming he was unaware that destroying the cargo would have this effect, the question *why* he ran the insurance company out of business would truly be "refused application." Unlike his answer

---

[21] *Id.*

to the question why he killed the passengers, the insurance bomber might answer *this* question with an expression of surprise—"Did I do *that*?". This gives us a basis for thinking that the *Why?* question test differentiates foreseen from unforeseen effects of an agent's action.

Recall, however, that a standard philosopher of action, like Hornsby or Davidson, thinks it unnecessary to distinguish foreseen from unforeseen effects in action-theoretic terms. Indeed, they might argue that if we are searching for an action-theoretic concept to which we can tie responsibility, we should look to the category of everything an agent *does*. Davidson suggests as much when he says that "[e]vent causality can spread responsibility for an action to the consequences of the action . . . ."[22] So instead of seeking to ground responsibility in the narrow class of *things done for a reason*, we might look to the wider category, namely the class of *all things agents do*. On this suggestion, then, the realm of responsible agency would extend past the realm of culpable (or praiseworthy) agency.

The difficulty with this suggestion is that while we previously could not account for responsibility judgments because the class to which they attached was too narrow, we now cannot account for them because the class to which they would attach is too broad, namely the class of all things agents do. Human beings are not responsible for everything of which they are agents in Davidson's sense. We want to be able to say the insurance bomber is responsible for killing the passengers, but not for running the insurance company out of business. But attaching responsibility to causal contribution would not allow us to distinguish the two.

Responsibility seems to occupy a middle category between things agents do for the sake of an end and all things they do. To provide a foundation for judgments of responsibility, we need some way of identifying this middle category in action-theoretic terms. The solution that rejects Thesis (2) in favor of a broader account of acting for a reason achieves this. The second solution we considered, that which rejects Thesis (1), may seem to identify the relevant middle category as well, since it retains the connection between responsibility and what an agent does intentionally. But it is ultimately unhelpful, insofar as it leaves the notion of the intentional itself unaccounted for. It thus fails to supply the internalist foundation we earlier claimed was desirable.

A thoroughly internalist account, by contrast, must explain intentional action in terms of a broad notion of acting for a reason. The category of things for which an agent is responsible is then the category of things done for reasons. In this Part I shall explore the idea that the notion of *choice* might ground an internalist account of the relation between the ends an agent sets for himself and the foreseen effects of his pursuit of those ends. Because foreseen side effects are themselves chosen, an account based on choice will allow us to distinguish *killing the passengers* from *running the insurance company out of business* in the way we have thought desirable. This would allow us to distinguish internal,

---

[22] Donald Davidson, *Agency*, *in* ESSAYS ON ACTIONS AND EVENTS 43, 49 (1980).

intentionalistic explanation from external, non-agentive explanation without limiting this distinction to that between things an agent does for the sake of an end, on the one hand, and everything else he does, on the other.

This solution involves two separate claims:

Thesis (4): If someone φ-s pursuant to a choice to φ, he φ-s *for a reason*.

Thesis (5): If someone foresees that in ψ-ing he will φ, then in ψ-ing he chooses to φ.

If Theses (4) and (5) are true, an agent does *for a reason* anything that he does, foreseeing he will do it. Since the insurance bomber foresaw the passengers' deaths, he killed the passengers for a reason, namely that he wanted the insurance money on the cargo. And if an agent does *for a reason* anything he foresees he will do in the course of acting, then he does all such things intentionally. If Thesis (3) is correct, then he is responsible for anything he foresees he will do, and he is not responsible for anything unforeseen in the course of acting. Thus, unlike Bratman, who would exclude a number of foreseen side effects from the ambit of what is done intentionally, I would include all fully foreseen effects. (We will discuss the question of partially foreseen effects in the next Part.) And I would exclude all unforeseen effects from the scope of responsible agency. What, however, is the argument for Theses (4) and (5)?

First consider Thesis (4). Thesis (4) asserts not merely that if an agent chooses to φ, and then φ-s, there is some description of his action under which he φ-s for a reason. This is true, but unhelpful in expanding the concept of acting for a reason. It also asserts that a person who chooses to φ, and actually φ-s, acts for a reason under the description of his action which is his φ-ing. Thus, the thesis asserts that if the insurance bomber chooses to kill the passengers, he kills them *for a reason*, even though he did not particularly desire their deaths. What could make this true?

When an agent chooses to accept a certain consequence of an action he performs, he is in some way endorsing that consequence, at least in relation to his chosen end. That is, the agent could decide to abandon his end when he sees that certain consequences will follow from the means he must adopt to accomplish it. The fact that he continues to pursue the end, given the consequences of doing so, suggests an important connection between the consequences an agent considers in selecting that end and the end itself. The consequences of the agent's action (both side effects and intended effects) thus stand in a certain *relation of value* to one another in the agent's deliberations, by which I mean that there is a value, or set of values, that explains the agent's willingness to do this-for-the-sake-of-that, a background system that relates the thing done as means to the thing aimed at as end.

Only given this background normative system can citing the agent's end explain the thing done: The rational explanatory force of the agent's end stems

from the fact that its relation to the thing done reflects some antecedent value the agent holds. From this it follows that an agent must have actually deliberated upon a consequence of an action in order for us to explain what he did as a product of his reason. It is explanation in light of an agent's reasons that allows us to see the agent as acting *for a reason*, and hence as acting intentionally. In the absence of this reflection, the thing the agent does is accidental from the standpoint of value—the agent cannot himself have endorsed it. Deliberation puts the stamp of rational agency on what an agent does.

What, now, about Thesis (5)? This thesis seems to require little argument. The suggestion is simply that an agent exercises choice over the effects of his actions he foresees. An agent need not choose something for its own sake in order to have chosen it. This way of thinking about choice seems to square with our ordinary use of the term. Imagine once again our outrage at the insurance bomber if he insisted that he had "no choice" but to kill the passengers, given that he had the end of destroying the plane. The fact that he could have avoided killing them by abandoning his plan, but chose to continue anyway, seems an adequate basis for saying he chose to kill the passengers. Choice, unlike intention, is non-teleological.

Together, Theses (4) and (5) suggest that an agent can be thought of as performing an action *for a reason* under any description under which it was foreseen by him. An action is intentional under any description under which it is foreseen, and hence it becomes plausible for us to say that an agent is responsible for his action under any description under which it was done intentionally. We can thus retain Theses (1) and (3) and reject Thesis (2). In this way, our account is fully internalist about responsibility.

VI.

What conclusions can we draw for criminal liability from the foregoing account of responsibility? A defendant who performs a prohibited act or brings about a prohibited result knowingly does so *for a reason*, and hence he does so intentionally. He is thus responsible for it in the descriptive sense we identified in Part II, and can be morally and legally evaluated for his behavior. By contrast, a defendant who is wholly unaware of the prohibited act does not do the prohibited thing for any reason. He is therefore not responsible for the prohibited act or consequence, and cannot be blamed for having done so. It would follow that there should be no criminal liability for negligent violations of a criminal norm.

Now even if one accepts the account I have offered, matters may not be this straightforward. For we have yet to consider cases in between the above two extremes, where the defendant is aware he is running a *risk* of violating a criminal prohibition. If the risk eventuates, does the defendant engage in the prohibited conduct or bring about the prohibited result *for a reason*? He certainly runs the relevant risk for a reason, since he knowingly does *that*. But this does not by itself tell us whether he brings the prohibited conduct or result about for a reason, in

virtue of having foreseen that he might bring it about. The account I have offered suggests that he does in cases in which the likelihoods are sufficiently high of his bringing it about, and otherwise not. If he thinks his act is highly likely to result in the death of another, for example, there is little difference between his responsibility in that case and the cases we have been considering in which he is certain to bring about death. At the other extreme, however, if the defendant is aware of a risk that a certain result will occur, but that risk is so minimal as to be almost non-existent, he approaches the agent who did not foresee the risk and is better thought of as not bringing about that result intentionally. He is probably not responsible for it in that case.

To be sure, there will be gray-area cases in the middle. Does a defendant who runs a fifty percent risk of killing someone kill that person intentionally? And if so, what about a slightly smaller or slightly greater risk? The uncertain response of the present account to such cases does not seem to count against it, given that such cases are unsettled both in morality and in law. The important point is that the cases in which we are inclined to treat the agent as responsible for the consequence will be ones in which we are also inclined to think he acted intentionally, whereas the ones in which we judge him not responsible are cases in which we think of his conduct as unintentional.

Notice that this approach to agents who knowingly engage in risky behavior comports with the criminal law's standard definition of recklessness. The Model Penal Code (MPC) defines recklessness as when a person "consciously disregards a substantial and unjustifiable risk" that the prohibited result will occur.[23] By requiring that the risk the defendant perceived be "substantial," the MPC exempts the defendant who brings about a highly unlikely harm from the ambit of recklessness.[24] The MPC does not indicate what level of likelihood is sufficient to count as "substantial;" a lacuna the Commentaries to the Code specifically endorse by saying that the matter is properly left to the jury's discretion.[25] But if "substantial" is a descriptive, rather than a normative concept, then the MPC's notion of recklessness roughly covers the same territory as the account of responsibility defended here.[26]

An interesting question arises in the case in which the defendant is aware of running a risk, but in which he underestimates the magnitude of the risk. The MPC does not indicate whether the substantiality of the risk should be included

---

[23] MODEL PENAL CODE § 2.02(c) (1962).

[24] Such a defendant would not even be "negligent" under the MPC's definition (§ 2.02(d)), since even the risk involved with negligence must be "substantial." The difference between recklessness and negligence is of course that the risk need not be perceived where the latter is concerned.

[25] MODEL PENAL CODE § 2.02 cmt. at 237 (1985).

[26] Peter Arenella has recently argued to me that the MPC's understanding of "substantial" is in fact normative rather than descriptive. And if this is correct, then recklessness as defined under the MPC would not be suited to a descriptive conception of responsibility.

within the agent's conscious disregard: Should the defendant be considered reckless as long as he was aware of running a risk, and the risk was substantial and unjustifiable, even if *his* perception of the risk was that it was not substantial? The better account, and one supported by the approach proposed here as well, is that the agent who consciously disregards a substantial risk is responsible for the consequence that eventuates from that risk only if he was aware not only of the risk's existence, but also of its substantiality. The defendant who believes a risk to be considerably smaller than it in fact is, such that he would not be responsible if the risks were as he supposed, would not be responsible for the consequence if it eventuates from the risk. But the reverse is not true. A defendant who thought a risk substantial when it was not would not be responsible for the eventuation of the risk, despite the fact that he believed he was acting in the face of a high likelihood of bringing about the very harm that did occur.[27]

It is worth noting an aspect of the MPC's approach to recklessness with which we might disagree. The MPC builds "justifiability" into the very definition of recklessness. The purpose of this provision is clear: the surgeon who performs a highly risky operation is not reckless, since he runs the risk of bringing about the patient's death justifiably. I believe this is a mistake. It conflicts with the idea of a non-moralized conception of responsibility, according to which responsibility is only the precondition for praising or blaming an agent for something he did. Saying he is not reckless is tantamount to saying he is not responsible for killing the patient. But that seems wrong. The surgeon who performs a highly risky operation is fully responsible for the death of his patient, if such results. But he is not to blame for having killed him, since he has a justification for having done so, namely that the patient stood to benefit from the operation, and consented to its occurrence on that basis. The MPC approach to recklessness builds the justificatory condition into the definition of the mental state, thus confounding prima facie conditions of responsibility with ultimate conditions, in just the way we saw with Hart's account above.

Accounting for responsibility in terms of foresight of harm, as Turner originally proposed, thus gives us a plausible approach to responsibility, an approach that easily generalizes to responsibility judgments in the criminal arena. It also makes sense of several important aspects of existing criminal law doctrine. In particular, it suggests that criminal negligence should be a much disfavored form of liability. It would place criminal negligence in a class of rather marginal doctrines of responsibility, such as complicit and vicarious liability, doctrines for which we have no justification other than their overall utility. But if, as I suggested at the outset, criminal responsibility is simply another form of

---

[27] Such cases are unlikely to arise in practice, particularly as there is a tendency to exaggerate the *ex ante* risks of an unlikely event occurring once it actually does occur. But if the defendant could convincingly show that the risk of which he was aware was in fact minute, he should be treated like the person who commits an attempt: he is only liable for any crime that attaches to believing he is behaving badly. He was not in fact behaving badly.

responsibility in ordinary morality, consequentialist arguments for its imposition should provide inadequate justification.

VII.

I shall conclude by considering two important objections to the account I have traced. The first objection is that the descriptive account of responsibility must presuppose some account of causation, since an agent must cause the conduct or result he foresees in order to be responsible for it. But it is commonly thought that the commonsense notion of cause already has the notion of responsibility built into it, and thus the former cannot be used to elucidate the latter. As William Dray has pointed out, if two historians were to debate "whether it was Hitler's invasion of Poland or Chamberlain's pledge to defend it which caused the outbreak of the Second World War," they *must* be discussing who was at fault for the outbreak of the war.[28] Or consider two cars that collide at an intersection, where one driver had a stop sign and the other a clear right of way. Our judgment that the driver who failed to stop at the sign caused the accident would surely reflect a judgment that the accident was that driver's fault for failing to stop. Otherwise we might as well say the accident was caused by the non-faulty driver's failing to drive just a little bit faster or a little bit slower. Cases such as these have led some authors to conclude that the notion of causation must be preceded by a moral theory, and that we cannot identify *anything* as the cause of anything else non-normatively.[29]

If the above claims about causation were correct, the account I have offered would be problematic. For if the notion of cause is parasitic on that of fault, then "being the cause of" cannot be an ingredient in a purely descriptive account of responsibility. Instead, responsibility must be imputed, since the causal relations on which it depends would themselves be imputed. If this is correct, there is no reason to resist responsibility in the absence of foresight of consequences. For we would have no reason not to "impute" responsibility to agents whose conduct falls below a certain normative standard. The appropriateness of such imputations would be a matter of the purposes we had in making them, as Hart early on suggested. And given that even an agent's unintentional behavior can be influenced or deterred with threat of sanction, punishment for unintentional conduct might suit our purposes.[30]

But one should not infer from the indeterminate nature of causation in *some* cases that causation is indeterminate everywhere. In particular, the central cases of responsibility we might consider display no causal indeterminacy. If I kill you by shooting you, no one would seriously deny that the shooting was the cause of your

---

[28] WILLIAM H. DRAY, LAWS AND EXPLANATION IN HISTORY 100 (1957).

[29] Arthur Ripstein, *Equality, Luck and Responsibility*, 23 PHIL. & PUB. AFF. 3 (1994).

[30] Indeed, it might turn out to be the case that imputing responsibility in the absence of causation suited our purposes. On this view, there would be no reason to resist such imputations either.

death, despite the fact that, like the discussion of Dray's historians, there are other causal *factors* required for my shooting to cause your death. As Joel Feinberg writes, "[e]xplanatory citations single out abnormal interferences with the normal course of events or hitherto unknown missing links in a person's understanding. They are designed simply to remove puzzlement by citing the causal factor that can shed the most light."[31] That we are not presupposing a judgment of culpability when we identify one causal factor as *the* cause is moreover evident from the fact that we can identify causal factors where human agents are not involved: the cause of that tree's falling over can be identified in terms of the "abnormal interference" of high winds.[32]

There are, however, two sorts of cases in which the normative account of causation seems particularly difficult to avoid. The first is one in which several causal factors vie for position as overall cause of an event, and there is no clear irregularity that suggests itself as cause. But such cases need not stand as a challenge to the descriptive account of responsibility. For on a descriptive account, it is acceptable to conclude that more than one agent is "responsible" for an occurrence, since this does not entail that more than one agent is to blame. Thus, Chamberlain might be "responsible" for the outbreak of war, along with Hitler, since actions of both contributed causally to that outcome, but one need not therefore conclude that both agents are to blame. If the supposed causal indeterminacy rests on the existence of multiple causal factors, it need not threaten our ability to offer a descriptive account of responsibility, since that account requires only the existence of a causal factor, not necessarily a unique causal relationship.

In this sense, judgments of responsibility are different from judgments of culpability. If we are asking who was to *blame* for a certain occurrence, we are normally asking for a unique identification, or at any rate, if more than one person was to blame, each must independently satisfy the moral criteria for blameworthiness. But where non-normative responsibility is concerned, it is not objectionable to ascribe partial responsibility to each causal factor, tailoring responsibility to degree of causal contribution. This common-sense approach to causation is only possible once responsibility is uncoupled from blame.

The second sort of case in which the normative approach may seem inevitable is one in which there is a clear irregularity to use as a basis for choosing among

---

[31] Feinberg calls this "the lantern criterion" for causation. Joel Feinberg, *Sua Culpa*, *in* DOING AND DESERVING: ESSAYS IN THE THEORY OF RESPONSIBILITY 204 (1970).

[32] Feinberg's proposal is subject to the objection that there are many instances in which we can inquire into the causes of ordinary events. Thus, we can ask what causes the tides, or the earth to turn, or the sky to appear red at sunset, even though these are the most ordinary of events. But we might distinguish here between events generally and particular events. Feinberg's suggestion seems correct as applied to the latter. Thus, we cannot easily ask what made the sky appear red at the end of the day *today*, unless it was unusually red, i.e., displayed an irregularity. The question otherwise can only be a request for an explanation of sunsets generally, and it is here that Feinberg's suggestion seems less helpful.

causal factors, but in which the irregularity can only be identified normatively. It is this way with omissions, as in the case of the driver who failed to stop at the stop sign. While these cases seem more challenging for the descriptive approach to responsibility, the following answer suggests itself. One need not infer from the fact that the irregularity is normatively identified that the causal relation is *itself* a normative one. That is, the law establishes a duty to stop at stop signs, in light of which other drivers come to have both a (normative) right to rely on the observance of this duty and a (descriptive) expectation that drivers will stop at stop signs. Where a driver fails to stop at a stop sign, the expected regularity is not present, and the breach of regularity can be identified as the cause of the resulting accident. That we can identify the driver who fails to stop at the sign as the cause of the accident without relying on a judgment of fault can be seen by considering the justified omission, namely the man who is driving his wife in labor to the hospital. We can still identify the failure to stop at the stop sign as the cause of the accident, even though the defendant has a justification for failing to stop and hence is not at fault. And this shows that the notion of cause can provide an independent element in a conception of responsibility, even where omissions are concerned.

A second objection to the account I have offered concerns my assumption that criminal responsibility mirrors our responsibility practices in ordinary morality. Given that individuals are sometimes held criminally liable for harm they did not foresee, and even sometimes for harm they did not cause, why do I think that criminal responsibility mirrors moral responsibility? Corporate responsibility and vicarious liability might further support one's suspicion that the criminal law is just a series of conventions, united under a single institution. The institution of criminal justice presumably is meant to satisfy some set of social purposes, just as the various institutions of tort law serve distinct social goals. On this view, the only constraint on forms of responsibility-ascription should be the contribution they make to the relevant institution's goals.

While the foregoing would constitute a perfectly coherent view of the criminal law, I do not believe it is our view. It is true that there are isolated instances in which criminal responsibility is merely "imputed" without regard for the underlying "facts" about responsible agency, but most rules of criminal liability nevertheless track our moral practices quite closely. At the very least, the core prohibitory norms of the special part, those based on common law crimes, are rules of moral disapprobation. And while there are many more recent regulatory offenses that cannot be readily intuited from our moral practices in the way that the common law crimes can, most of these are reasonable extensions of common law crimes. Those that are not might be questioned.

To be sure, we could conceive the criminal law along different, more utilitarian lines. We could regard criminal law exclusively as an instrument for social control, and treat the rules of criminal liability as a list of prices meant to discourage inefficient behavior. But, as many theorists have noted, the institution

would probably lose its effectiveness in this form.[33] The criminal law's connection with morality is a crucial part of its ability to deter and to contribute to the teaching of moral standards. There is thus a consequentialist argument for designing rules of criminal conduct that mirror the practices of responsibility we follow in our moral lives, that is, a consequentialist reason for maintaining an institution that incorporates norms of responsibility and blame.

---

[33] Paul Robinson & John Darley, *The Utility of Desert*, 91 NW. U. L. REV. 453 (1997).