

Harm, History, and Counterfactuals

STEPHEN PERRY*

TABLE OF CONTENTS

I.	INTRODUCTION	1283
II.	HARM, COUNTERFACTUALS, AND HISTORICAL WORSENING	1285
III.	SHIFFRIN'S CRITIQUE OF THE COMPARATIVE MODEL.....	1299
IV.	HARM, INTERESTS, AND RIGHTS.....	1304
V.	HARM AND COMPENSATION	1309

I. INTRODUCTION

In this Article, I undertake a very preliminary inquiry into some aspects of the concept of harm. My excuse for doing so in a symposium on compensation is that, in private law and particularly in tort law, an award of damages is often intended to compensate for harm; if we do not know something about the nature of harm, we cannot fully understand the nature of at least this type of compensation. To avoid one possible source of confusion, I should add immediately that harm is not the only thing that can be compensated by an award of compensatory damages. Compensation in law is generally meant to rectify a setback to an

* Fiorello La Guardia Professor of Law and Professor of Philosophy, New York University School of Law. I am grateful for very insightful responses to an earlier version of this Article by John Goldberg at the symposium on compensation held at the University of San Diego in February 2003 and by Andrew Simester at the Oxford Jurisprudence Colloquium held at Oxford in March 2003. I also benefited a great deal from the comments of other participants on both occasions. Because I was able to revise the Article for publication in only the most limited way, I have unfortunately not been able to discuss many of the points raised by Goldberg, Simester, and others. I hope to do so in future work.

interest, but, I shall argue, while all instances of harm are setbacks to interests, not all setbacks to interests are instances of harm. Further, in cases where the law imposes liability for omissions—that is, for breaching an affirmative duty to put someone in a certain position—it may be that the term “compensation” is appropriate even if the person being compensated has not suffered any setback, in the sense of an historical worsening, at all. However, nothing that I have to say about harm will turn on accepting one account rather than another of the concept of compensation.

To avoid one other possible source of confusion, I should also emphasize at the outset that my inquiry is into the nature of harm and not into the nature of causation. I shall argue that an instance of harm is not simply a certain kind of condition or state of being, but rather must have been brought about by a causal process. While causation and harm are obviously distinct concepts, it might be thought that, because harm is the result of a causal process, one’s analysis of causation must carry over to one’s analysis of harm. If causation is properly analyzable in counterfactual terms, for example, it might be thought that the same must be true of harm.¹ That, however, would be a mistake. To ask what it means to say that event *A* caused event *B* is one question; to ask what it means to say that event *B* is harmful is another.

In Part II, taking a seminal article by Joel Feinberg as my starting point,² I consider two possible characterizations of harm. Both treat harm as a setback to an interest and as involving a comparison between two states of affairs, but one supposes that the relevant comparison is based on a counterfactual inquiry, while the other supposes that it is based on an historical inquiry. I argue that the historical characterization is the preferable one. In Part III, I consider a critique offered by Seana Shiffrin of comparative accounts of harm in general.³ In Part IV, I try to say something further about the relationship among harm, interests, and rights. Finally, in Part V, I present some considerations in favor of the view that even though harm should be characterized historically, compensation for harm in tort should often be determined counterfactually, by reference to what would have happened had the tort not occurred.

1. Michael Moore very persuasively criticizes the counterfactual analysis of causation in Michael Moore, *For What Must We Pay? Causation and Counterfactual Baselines*, 40 SAN DIEGO L. REV. 1181 (2003).

2. JOEL FEINBERG, *Wrongful Life and the Counterfactual Element in Harming*, in FREEDOM AND FULFILLMENT 3 (1992).

3. Seana Valentine Shiffrin, *Wrongful Life, Procreative Responsibility, and the Significance of Harm*, 5 LEGAL THEORY 117 (1999).

II. HARM, COUNTERFACTUALS, AND HISTORICAL WORSENING

Speaking of “harm” in the sense that he takes to be presupposed by the harm principle and, more generally, to be of interest to the civil and criminal law, Joel Feinberg offers the following analysis:

A harms *B* in the relevant sense if and only if:

1. *A* acts (in a sense wide enough to include omissions and extended sequences of activity).
2. *A*'s action is defective or faulty with respect to the risks it creates to *B*, that is, it is done either with the intention of producing the consequences for *B* that follow, or similarly adverse ones, or with negligence or recklessness in respect to those consequences.
3. *A*'s acting in that manner is indefensible, that is, neither excusable nor justifiable.
4. *A*'s action is the cause of an adverse effect on *B*'s self-interest (a “state of harm”).
5. *A*'s action is also a violation of *B*'s right.⁴

Feinberg notes in a footnote that condition 5 might, given conditions 3 and 4, be redundant. A more important point, however, concerns whether or not conditions 2, 3, and 5 are appropriately included as necessary conditions at all in the analysis. Feinberg is clearly analyzing the notion of a harm in moral terms, and he is right to do so. This is true if for no other reason than that condition 4, which I take to be the heart of the analysis, requires that *A*'s act have an adverse effect on *B*'s self-interest. Determining what constitutes a person's self-interest, as well as determining what effects on self-interest should count as adverse, clearly involves substantive moral argument. However, the fact that the analysis of “*A* harms *B*” is at least partially moral in character does not entail that *A*'s harming act must have been, say, faulty or defective (condition 2). Even if we follow Feinberg and restrict our attention to harms that might be of concern to the law, insisting on conditions 2, 3, and 5 clearly leaves out of account the possibility of certain theoretical approaches to tort law in particular; condition 2, for example, is incompatible with strict liability. Perhaps strict liability is not in the end a defensible theory of recovery in tort, but that is a substantive conclusion, and it needs to be defended in substantive terms; in the absence of such argument, Feinberg's analysis begs an important question. Intuitively, it seems clear that if *A* acts and thereby adversely affects *B*'s self-interest, she has harmed *B*, whether or not she acted faultily or violated one of

4. FEINBERG, *supra* note 2, at 6.

B's rights. Fault or a rights violation may well be a necessary condition of moral blameworthiness on *A*'s part, or of her civil or criminal liability at law, but there does not seem to be any basis for saying that in their absence *A* did not, in fact, harm *B*.

In this Article, I shall therefore focus on Feinberg's conditions 1 and 4. It should be noted that condition 4 speaks of an adverse effect on "*B*'s self-interest," as though *B* has just one interest at stake. The reason Feinberg phrases the point this way would appear to be the following. Although he states earlier in his article that "[t]he term 'interests' is best left undefined here, except to say that interests are distinguishable components of a person's good or well-being,"⁵ he immediately goes on to say that "our concepts seem to commit us to the view that interests can be summed up or integrated into one emergent personal interest."⁶ This last point, however, is far from evident, and in Part IV, I shall argue that it is mistaken. Condition 2 should therefore be reformulated to speak simply of an adverse effect on "one of *B*'s interests" rather than on "*B*'s self-interest."

This brings us to a crucial point, which is that, in Feinberg's view, the analysis of "*A* harms *B*" as thus far presented is incomplete. We also need, as an independent requirement for the occurrence of harm, the so-called counterfactual condition:

6. *B*'s personal interest is in a worse condition (usually but not always lower on the interest graph) than it would be in had *A* not acted as he did.⁷

Feinberg distinguishes the counterfactual test from the worsening test, which is not, he says, always required for an act of harming. The worsening test can be expressed as follows:

- 6X. *B*'s personal interest is in a worse condition (lower on the interest graph) than it was before *A* acted.⁸

For purposes of clarity, I will refer to what Feinberg calls the worsening test as the "historical worsening test." As Feinberg notes, the counterfactual test and the historical worsening test do not amount to the same thing. Feinberg does not use this kind of example, but consider an instance of concurrent causation, such as occurred in the famous case of *Kingston v. Chicago & Northwestern Railway Co.*:⁹ Two separately started fires joined and burned down the plaintiff's house, where either

5. *Id.* at 4.

6. *Id.*

7. *Id.* at 7.

8. *Id.*

9. 211 N.W. 913 (Wis. 1927).

fire would have been sufficient by itself to bring about that result.¹⁰ Suppose *A* started one of the fires. Assuming (as seems inevitable) that both fires are causes of the house burning down, *A* causally contributed to the historical worsening of *B*'s interest in his house. (For present purposes, we do not need to specify exactly what the nature of that interest is.) The historical worsening test is thus satisfied. But the counterfactual test is not satisfied, because the house would have been burned down anyway by the other fire. Since Feinberg makes the counterfactual test rather than the historical worsening test the touchstone of harm, it would seem that, on Feinberg's view, *A* has not harmed *B*.

Feinberg uses the following example to show that the counterfactual test can be satisfied even when the historical worsening test is not. Imagine that *B* is a model who is wrongfully detained by *A* and thereby prevented from taking part in the Miss America Pageant. If she had been able to enter the contest, *B* would have won a million dollars. Feinberg holds that *B*'s condition has not been historically worsened, "because she is no worse off than before the detention."¹¹ But because she is worse off than she would have been if she had not been detained, the counterfactual test is satisfied, and she is harmed. Feinberg in fact argues that sometimes *A* can harm *B* even though *A* not only did not historically worsen *B*'s condition but actually bettered it. Feinberg asks us to imagine that doctor *A* treats *B*, her patient, and improves his condition, but because she was negligent she did not improve it as much as she should have done. Although *A* bettered *B*'s condition in an historical sense, Feinberg claims that *A* harmed *B* because the counterfactual condition has been violated: *B* is worse off than he would have been had *A* behaved as she ought to have done (as she had a duty to do). As Feinberg remarks in a footnote, this example only shows what he wants it to show if condition 6 is revised to hold that *B*'s condition is in a worse state than it would have been in if *A* had acted as she should have done, rather than if she had not acted as she did. I will return to this kind of case at the end of this Part.

In light of the above considerations, let me reformulate Feinberg's analysis of *A*'s harming *B* along the following lines:

10. *Id.* at 914.

11. FEINBERG, *supra* note 2, at 7.

- (1) *A* acts (in a sense wide enough to include omissions and extended sequences of activity).
- (2) *A*'s action is a cause of an adverse effect on one of *B*'s interests (a "state of harm").
- (3) *B*'s personal interest is in a worse condition (usually but not always lower on the interest graph) than it would be had *A* not acted as he did ("the counterfactual condition").

Floating in the background somewhere, but apparently not a part of the analysis, we have:

- (3X) *B*'s personal interest is in a worse condition (lower on the interest graph) than it was in before *A* acted ("the worsening condition").

I mention, in passing, that it is not entirely clear what the relationship is supposed to be between conditions 2 and 3. Does condition 2 require an actual historical worsening, where we must then look to condition 3 to determine if that worsening was "harm"? This does not seem consistent with Feinberg's understanding of the Miss America example (or, for that matter, with the words "a state of harm" that he places in parentheses to describe condition 2). Probably the better view is that conditions 3 and 3X are meant to be alternative interpretations of condition 2, where Feinberg himself favors condition 3 as the preferable interpretation.

Henceforth, when I refer to Feinberg's analysis of harming, it will be to this reformulated version, involving conditions 1 through 3.

Feinberg considers a possible problem to which the counterfactual test might be thought to give rise. The standard hypothetical of the doomed airplane helps to make this problem clear. *B* gets in *A*'s taxicab to be driven to the airport. En route, *A* drives negligently and hits another car. *B* breaks his leg in the accident and as a result misses his flight. The plane that *B* missed crashes soon after takeoff, and all aboard are killed. It is as certain as anything can be that had *B* caught the flight, he too would have been killed. Feinberg asks if *A* harmed *B* by negligently causing his leg to be broken. He concedes that while common sense seems to say yes, his analysis appears to say no. Applying the counterfactual test, *B*'s overall interests would not only have been not better served on balance had *A* driven his cab carefully, they would in fact have been worsened (because *B* would have been dead).

Feinberg considers the possibility that the way to deal with this and related difficulties might be to replace condition 3 in his analysis, which is the counterfactual test, with a disjunctive condition consisting of conditions 3 and 3X. Thus *A* harms *B*, as a result of an act by *A*, only if *either* *B*'s condition is historically worsened by the act *or* *B* is worse off

than he would have been had *A* not acted. But Feinberg does not think that this disjunctive solution will work either. Suppose, he suggests, that *A*'s gang abducts *B* the day before the Miss America Pageant, thus preventing *B* from participating and thereby winning a million dollars. However, if *A*'s gang had not abducted her, then *C*'s gang, in a completely independent operation, would have done so. Feinberg argues that the historical worsening test is not met for the reason that he gave in the earlier variant of this example, namely, because *B* is no worse off historically after the abduction than she was before. But the counterfactual condition is also not met: *B* is not worse off than she would have been had *A* not abducted her, because if *A* had not done so, then *C* would have. Feinberg thinks it is nonetheless intuitively obvious that *B* was harmed in this example.

Although he is not entirely clear about this, Feinberg's response to these difficulties is apparently to reject the disjunctive solution and to retain the counterfactual condition, namely condition 3, as the touchstone of when one person has caused another person harm. Feinberg believes that the doomed airplane case and the second Miss America case have a common structure. A first harmer, *H*₁, puts *V* in a worse position for a period of time *t*. After *t*, a second harmer, *H*₂ would have caused equal or greater harm. *H*₁ is *V*'s actual harmer, "partly in virtue of his satisfying the counterfactual condition for time *t*."¹² But, Feinberg asks, consider the case in which *V*'s injuries linger on after *t*; for example, the would-be airline passenger's leg is still broken even after the plane has crashed. Feinberg assumes that the law would nonetheless permit the passenger to recover compensation even for the injury after time *t*, and comments as follows:

The law, I think, implicitly assumes a distinction between the harm actually caused to *V* by *H*₁ on balance (which is determined in part by the counterfactual test) and the harm that *H*₁ is justly answerable for. The harmful state that *V* remains in is a kind of "causal residue" directly resulting from the harm that *H*₁ *did* cause. It is unquestionably a *harmful* condition but strictly speaking after the passage of interval *t*, it is not a *harmed* condition, that is, not a net harm produced by *H*₁'s act of harming. Nevertheless, it is perfectly just to hold *H*₁ liable for the full damages, including those beyond what he actually caused, because only an unforeseeable fluke of chance (e.g., a plane crash) accounts for the brevity of the period during which the counterfactual condition continued to be satisfied. In short, *H*₁ harmed *V* because there was some period, however brief, during which the counterfactual condition was satisfied, and he is justly answerable for the

12. *Id.* at 10.

effects of the harm he originally caused, since it was only a fluke of chance, not anything he can claim credit for, that limited the scope of his harming.¹³

Feinberg may or may not be right about the justice of H_1 's compensating V . That is a question to which I shall return in Part V. But I believe the details of his account are nonetheless mistaken. As an aside, I note that it is very difficult to see how Feinberg would apply his own solution to the variant Miss America problem, because it is not clear that, by Feinberg's own lights, B was worse off, between the time of A 's abduction and the time that C 's abduction would have taken place, than she would have been if A had not abducted her.¹⁴ For the moment, however, let us leave Miss America to one side and focus instead on the doomed airplane hypothetical. Feinberg might be taken to be suggesting, in the above passage, that H_1 did not *cause* the harmful condition from which V was still suffering at time t and following. That would be a very strong claim, and a very implausible one on any acceptable theory of causation. But how else are we to construe Feinberg's argument? Perhaps the clue resides in the claim that there was no *net* harm. This claim is presumably based on Feinberg's thesis, mentioned earlier, that all interests can be "summed up or integrated into one emergent personal interest." I shall argue in Part IV that there is no good reason to accept this thesis, but even if it were correct, the result would surely be that H_1 has not harmed V *at all*, not just that he has not harmed him after time t . The plus to be weighed against the minus of B 's broken leg is, after all, the saving of his life. So long as we accept a counterfactual rather than an historical worsening interpretation of harm (condition 3 rather than 3X), there does not seem to be any good reason to limit the "netting out" effect, if there is one, to time t and after. While it is true that there is a period between the time when the leg was broken and the time of the crash during which it would be possible to say that, applying the counterfactual criterion to that period only, V was worse off, it is not clear why we should limit the application of the test *during* this period *to* that period. Given that the counterfactual test is not an historical test, why should it not be applied on an all-in basis which takes account of everything that would have happened if H_1 had not acted as he did? Even at a time after the accident but before the crash, is V not better off, counterfactually speaking, than he would have been if H had not negligently caused the accident?

13. *Id.* at 10–11.

14. It is not entirely clear what Feinberg thinks the harm consists of in this example. The most plausible candidates are not being able to enter the pageant, being prevented from winning the pageant, and being prevented from collecting the million dollars. All of these events, we can assume, take place after the time that C 's abduction would have taken place.

Even if I am wrong to conclude that Feinberg's reliance on interest-summing and netting-out is, in each case, problematic, there are other difficulties with his counterfactual account of harm. Consider a variant on the *Kingston* case discussed above. Suppose that *A* sets a fire which burns down *B*'s house.¹⁵ There is another fire that would have burned down the house if *A*'s fire had not already done so, but it is prevented from having that effect because *A*'s fire consumed some of the fuel that lies between the second fire and the house. It turns out that, if *A*'s fire had not existed, the second fire would have reached *B*'s house at the exact moment when, in fact, *A*'s fire did. Thus, there is no period *t* during which it is possible to say that *B* was (counterfactually) worse off than if *A* had not set the fire. According to Feinberg's account, *A* did not harm *B*. But this seems plainly wrong. Perhaps *A* does not have to compensate *B* for the harm he caused, but that is another question. It seems indisputable that, in burning down *B*'s house, *A* caused *B* harm, and that this is true even if the same kind of harm, in the same degree, would have been caused at the exact same time by some other fire had *A*'s fire not existed.

I have concentrated so far on cases in which the counterfactual understanding of harm would lead us to say that there is no harm but where, intuitively, harm exists. But there are also many cases in which the counterfactual understanding would lead us to say that there is harm where, intuitively, there is none. This is because, as Shiffrin points out, a pure counterfactual account of harm does not, and cannot, draw any distinction between harming and failing to benefit.¹⁶ I am not harmed just because, were things different in some relevant way, I would have been better off than I am now.

The solution to all these various difficulties is, I think, the following. Instead of analyzing harm by reference to the counterfactual condition (condition 3), we should analyze it by reference to the historical worsening condition (condition 3X). Our concept of harm is such that harm only occurs if there is an actual worsening, that is, an historical setback to an interest: If *B* was harmed, then there must be some relevant interest of *B*'s that was initially at level *n* and that was caused by the

15. I note in passing that, for the reasons discussed earlier, the question of whether *A* harmed *B* does not seem to depend on whether or not *A* acted wrongfully in setting the fire, although of course the question of whether *A* owes *B* compensation might well so depend.

16. Shiffrin, *supra* note 3, at 121.

allegedly harmful event to be at a level below *n*. I should note immediately that I am putting this requirement forward as a *necessary* condition for the occurrence of harm. I am not suggesting that it is a *sufficient* condition.

Why, it might be asked, should we not adopt a conjunctive solution, according to which *A* harmed *B* only if *both* conditions 3 and 3X are met? Perhaps we might say, using the variant of *Kingston* just discussed, that the historical worsening caused by *A*'s fire was "protoharm," but because the counterfactual condition was not met, *B* did not suffer harm in the true or full sense. The answer to this suggestion has to be, I believe, that it simply does not conform to our ordinary concept of harm. Our ordinary concept is relatively minimalist. What I have called protoharm just *is* harm; no further adornment is required to turn it into the real thing. Of course, there are further interesting questions about responsibility and liability for harm caused, and the answers to these questions can be quite nuanced. For example, the law accepts that whether or not *A* owes *B* compensation in the variant *Kingston* hypothetical just discussed depends on whether or not the second fire was set tortiously. If it was set tortiously, then *A* owes compensation even though the counterfactual condition has apparently not been met.¹⁷ I will suggest in Part V that the law is correct on this point, and if that is so, then Feinberg's counterfactual analysis of harm would presumably have to be revised or qualified in some way. The main point for present purposes, however, is that whether or not the law is right here, it is surely a mistake to fold these difficult questions of moral and legal responsibility into the concept of harm itself. The paradigm of harm is an historical worsening. Once we have established that *A* caused *B* harm in this sense then further questions of responsibility and liability arise, and these might well involve counterfactual analysis. But these questions are distinct from the determination of whether or not harm occurred in the first place.¹⁸

What are we to say, then, of Feinberg's Miss America hypothetical? In both its versions (one abductor in the first variant, one abductor and another waiting in the wings, in the second), Feinberg maintains that *B*, the abductee, suffered no historical worsening; she is no worse off after the abduction than she was before. Feinberg thinks that she suffered harm in both variants because she is worse off counterfactually; if she had not been abducted, she would have won the contest and a million dollars. But Feinberg is wrong to suggest that there has been no

17. This is, at least, the law according to *Kingston* itself. *Kingston v. Chi. & N.W. Ry.*, 211 N.W. 913, 914–15 (Wis. 1927).

18. There are some limited exceptions to this generalization, such that the moral quality of the act leading to harm affects the character and extent of the harm itself. I will discuss such cases later in this Part.

historical worsening here. *A* deprived *B*, the would-be Miss America, of an option or an opportunity to enter the pageant, and such a deprivation is a setback to her autonomy of a kind that constitutes harm.¹⁹ The failure to win the contest and the million dollars are instances of consequential damage flowing from this initial harm. I do not think it matters in this regard that *A* violated a right of *B*'s (by wrongfully detaining her).²⁰ Suppose instead that he had innocently and nonnegligently misdirected her, telling her that the pageant was in Atlantic City when for some obscure and generally unknown reason it was being held that year in Houston. I believe the proper conclusion is that *A* (innocently) caused *B* harm, because, again, *A* interfered with *B*'s autonomy by depriving her of a valuable option or opportunity. He induced *B* to rely on the assumption that the pageant was in Atlantic City and, accordingly, to behave in a way that eliminated her option of getting to Houston, where the pageant was actually to take place. The loss of the relevant option flows from *B*'s detrimental reliance. It is possible that there are contextual limits on the manner in which *A* can affect *B*'s options and still be said to have harmed her. For example, suppose that *A* did not innocently direct *B* away from the pageant but instead cancelled the entire affair the year that *B* was hoping to enter. (*A* was the pageant's CEO, if there is such a person, and he acted justifiably, for reasons relating to the pageant's shaky finances.) *A* has still deprived *B* of the option of entering the pageant, but it is not entirely clear that *B* has been harmed here, at least in the way that she would

19. Cf. JOSEPH RAZ, *THE MORALITY OF FREEDOM* 413 (1986). In his comment on this Article, John Goldberg seems to assume that I am committed to treating "lost expectancies," meaning interferences with the expectation interest in contract law, as a subcategory of lost opportunities. See John C.P. Goldberg, *Harm, Injury, and Proximate Cause*, 40 *SAN DIEGO L. REV.* 1315, 1328–29 (2003). But this is a misunderstanding. In the case of a lost opportunity, *A* has done something (detained *B*, induced *B* to rely to her detriment, etc.) such that *B* no longer has an option to do something which she had before *A* acted. Interferences with the expectation interest in contract law can involve the loss of an opportunity in this sense, but they need not. Rather, an interference with the expectation interest constitutes the violation of a certain kind of right, and, as we shall see in Part IV, the violation of a right does not necessarily involve either an interference with autonomy (the deprivation of an option) or, indeed, harm of any kind. Conversely, interferences with autonomy do not necessarily amount to the violation of a right. Consider the case in the text in which *A* innocently misdirects *B* to Atlantic City, thereby depriving her of the option of getting to Houston and entering the pageant there. *A* has interfered with *B*'s autonomy, but it would not be plausible to say that he has violated one of her rights.

20. Such a detention is, of course, harm in itself, partly but not entirely because it constitutes an interference with autonomy.

have been harmed by being prevented from entering a contest that was in fact being held and that others were free to enter. This seems to be the case even if, as a counterfactual matter, *B* would have won the pageant had it in fact been held. But this is a speculative point, and nothing turns on it for subsequent purposes.

As we saw at the outset, Feinberg is concerned to analyze the notion of “*A* harming *B*” rather than the notion of “harm” as such. I have already suggested that his original conditions 2, 3, and 5, which were concerned with the moral character of *A*’s act (was it faulty, a rights violation, excused or justified, and so on), are irrelevant to the characterization of the outcome for *B* as harm. I now wish to suggest, along much the same lines, that the requirement of human agency in Feinberg’s conditions 1 and 2 (on the slimmed-down version of Feinberg’s analysis that I presented earlier) is similarly irrelevant. Condition 1 holds that *A* acted (or in a suitable sense omitted to act), while condition 2 holds that *A*’s action is the cause of an adverse effect on one of *B*’s interests. On the minimalist, historical analysis of harm that I have suggested lies at the core of our ordinary concept, the gist of harm is the historical worsening of one of *B*’s interests. Such a worsening is the result of a causal process; it is an event that has been precipitated by other events. In general, there is no reason to think that the analysis of harm changes in any significant way if one of the precipitating events was an instance of human agency.²¹ There is no reason, in other words, to think that the relational notion of “*A* harms *B*” is morally richer or more complicated than the conjunction of the concept of harm and the concept of agency. Harm is a moral concept, in the sense that we require moral argument to establish what is and is not harm, but I think the relevant moral argument is for the most part limited to *what* happens and does not extend to *how* it happens.

On the basis of the foregoing, I conclude that there is no distinct relational moral category of the kind Feinberg apparently sets out to describe. This means that there is no *general* moral category that takes the form “*A* harms *B*.” But it does not follow that there cannot be special kinds of harm that can only be caused by human actions. For example, it is plausible to think that the primary manner in which

21. Andrew Simester made the point in his comments at the Oxford Jurisprudence Colloquium that it seems odd to say that a person has been harmed by nature, as opposed to having been harmed by another person. It is no doubt true that the subject of the verb “to harm” is usually a phrase referring to a person, and perhaps it does seem somewhat stilted to say that a storm, say, harmed someone. But I think there are many other locutions that are perfectly natural and that make clear that it is not just human agency that can cause harm. I think it is quite appropriate, for example, to say that someone suffered harm during a hurricane.

dignitary interests are set back involves the deliberate actions of others. It is no doubt true that one's dignity can be adversely affected in many different ways, but there nonetheless seems to be a distinctive kind of harm that is suffered when one is subjected to deliberate indignity by another person. R.A. Duff has argued along similar lines that the primary harm associated with rape is internal to the act of rape: "[E]ssential to the description and identification of the harm as a harm is a human action which perpetrates it."²² The harm in such cases presumably consists partly in the interference with autonomy, bodily integrity, and emotional well-being, and partly in the fact that this state of affairs was brought about intentionally, knowingly, or recklessly by another human being.²³ The general point to be emphasized for present purposes, however, is that most harms are not like this. A broken leg is a broken leg, and this is true even if the leg was broken as a result of another person's agency.

The upshot of the discussion so far, then, is that the core characteristic of harm, and a necessary condition of the occurrence of harm, is the historical worsening condition:

HWC: A person has been harmed only if some relevant interest of that person has been affected adversely, meaning the interest has been caused to worsen or deteriorate in time.

Let me add three quick points of clarification. First, the HWC claims that all harms are historical setbacks to interests. It does not follow, however, that all historical setbacks to interests are harms. That is why I

22. R.A. Duff, *Intentions Legal and Philosophical*, 9 OXFORD J. LEGAL STUD. 76, 90 (1989); see also R.A. DUFF, INTENTION, AGENCY AND CRIMINAL LIABILITY: PHILOSOPHY OF ACTION AND THE CRIMINAL LAW 111–15 (1990) [hereinafter DUFF, CRIMINAL LIABILITY].

23. In John Goldberg's comment on this Article, he apparently takes me to be saying that any rights violation is a harm of this special kind. See Goldberg, *supra* note 19, at 1319. But this is a misunderstanding. In my view, this special kind of harm—to use Goldberg's helpful terminology, "relational" or "dyadic" harm, which is to be contrasted with the usual case of "monadic" harm, *see id.*—involves interference with a particular class of interests, of which dignity is the main (and perhaps only) instance. The harm in rape is relational rather than monadic because an interest of this sort was adversely affected (along with other interests, of course). Perhaps relational harms are usually rights violations, but the two categories do not coincide, and neither is a subset of the other. Note that while it is true that a right can only be violated by a person, it does not follow that the harm (if any) which is associated with a rights violation must be of a kind that can only be caused by human action. I further discuss the relationship between harm and rights in Part IV.

have formulated the HWC in terms of *relevant* interests rather than just interests. I will return to this point later. The second point is one that I mentioned at the beginning of this Article. While I have rejected the claim that harm should be understood by reference, in whole or in part, to the counterfactual condition, this does not mean that a counterfactual inquiry has no place in determining whether or not harm has occurred. Harm is the result of a causal process, and the correct analysis of causation might well require us to engage in counterfactual inquiry in order to determine what is a cause of what.²⁴ But even if it is true that we have to employ counterfactuals to characterize properly the causation of a harmful event, the determination that the event was a *harmful* one does not depend on counterfactual inquiry. It depends, rather, on an historical comparison of before and after. The third point is that even if a counterfactual analysis does not figure in the determination of whether an event was harmful, it might well figure, as has already been noted, in the quantification of damages in tort law. I discuss this issue in Part V.

As I have characterized the historical worsening account, it does not require a particular account of causation, but it does require that harm have been brought about by a causal process. I have further argued that the harmful character of the end result is determined historically rather than counterfactually. Now it is true that one could offer a mixed account of harm, by which I mean an account that requires harm to be the outcome of a causal process but that determines the harmful character of the end result by means of a counterfactual inquiry. As a general matter, however, this grouping of caused event and counterfactual characterization of harm seems quite artificial. It seems designed *simply* to get around the fact that a pure counterfactual analysis of harm—that is, one that compares the current state of affairs to what would or should have happened, without inquiring into how the current state of affairs came about—will be completely incapable, without more, of distinguishing a harm from a failure to benefit. The more natural correlation would seem to be between the requirement that there be a causal process and the characterization of the end result of that process as harmful because it was an historical worsening. This is more natural because an historical

24. As was noted earlier, Michael Moore makes a very strong case for the view that causation is not properly understood simply by reference to a counterfactual test. *See supra* note 1. In my view, the best analysis of causation is given by what Richard Wright has called the NESS test: A cause is a Necessary, i.e. nonredundant, Element of a Sufficient Set, meaning a set of conditions minimally sufficient to bring about the effect in question. *See* Richard W. Wright, *Causation in Tort Law*, 73 CAL. L. REV. 1735, 1774 (1985). This understanding of causation is based on the idea of instantiating a causal law. It does not reduce causation to a counterfactual test, but counterfactual inquiry will nonetheless be involved in determining whether a particular chain of events is or is not an instantiation of some causal law.

worsening just *is* a certain kind of causal process. The requirement of a causal process has not, as with the mixed account, been tacked on to an otherwise self-sufficient and freestanding account of harm, namely, the pure counterfactual approach, simply to avoid certain counterintuitive consequences.

In arguing for the historical worsening condition, I have pointed out what I take to be serious difficulties for an understanding of harm based on the counterfactual condition. This is not to say, of course, that an understanding based on the historical condition is without problems of its own. In the remainder of this Part, I would like to discuss some difficult cases that have been drawn to my attention by commentators. First, consider the case of a preconception tort, in which someone does something to the body of a woman who is likely to become a mother that has the effect, when she subsequently becomes pregnant, of preventing the fetus from developing two arms.²⁵ Intuitively this appears to be a case of harm, even though there is no obvious historical worsening. (The fetus never lost an arm; it only had one from the outset.) Consider also a case in which someone acts in such a way as to prevent someone

25. This hypothetical is based on, but different from, one posed by John Goldberg in his comment. See Goldberg, *supra* note 19, at 1329–30. In the Goldberg variation, a couple unreasonably fails to take a pill that, if taken within a few hours before or after sexual intercourse, would have prevented a birth defect in their child. I have not used Goldberg's hypothetical because it contains a complication, namely, the question of whether the birth defect is the result of an act or an omission, that somewhat obscures our intuitions. If the defect is properly described as the result of an omission, then I do not think it is appropriately described as harm. This is because pure omissions cannot, in my view, be causes, whereas harms are always the result of a causal process. On the question of omissions as causes, see Moore, *supra* note 1, at 1222–26; Michael S. Moore, *Causation and Responsibility*, SOC. PHIL. & POL'Y, Summer 1999, at 1, 23–24, 31–34. As Moore argues, the relations in causal statements are events, not facts, and while an omission is a fact, it is not an event. It is worth noting that if one adopted the contrary view, namely that omissions can be causes (at least in cases where the omitted act could have prevented harm or conferred a benefit), and combined it with a full-fledged counterfactual analysis of harm, the distinction between harming and failing to benefit would disappear completely, even on what I called in the text a mixed account of harm. This might not be an argument in itself against the conclusion that omissions can be causes, but it is an unsettling result. None of this is meant to deny, of course, that we can have affirmative duties to bring about certain states of affairs and that we can properly be labeled responsible for failing to bring about those states of affairs. Sometimes such failures will involve harm, as, for example, when the right holder relies to her detriment on the expectation that the other party will act in accordance with his duty. Often, however, the breach of an affirmative duty will simply amount to a failure to confer a benefit. See further *supra* note 19 and the discussion at the end of Part IV.

else from recovering from an injury or a disease.²⁶ Intuitively this too appears to be a case of harm, but again there seems to be no historical worsening.

The counterfactual approach could explain why there is harm in these cases, but I believe the general difficulties with that approach make it too unattractive a solution to adopt on even an ad hoc basis. I am far from certain what the correct analysis of such cases is, but I would like to suggest very tentatively that they may, despite initial appearances, in fact be instances of historical worsenings. The first point to be noted is that both cases involve a causal process. The second is that that process affects what might be called a natural potential. In the one case we have the natural potential and tendency of a fetus to develop, in the absence of external interference, two arms, and in the other we have the natural potential and tendency of the body to heal itself, again in the absence of external interference. If this potential is absent in either case, it seems clear that there is no harm. In both cases, the being who has been affected can be said to have an interest in the realization of the relevant potential. (I am assuming for present purposes that a fetus has interests.) I acknowledge that the idea that harm can take the form of interference with a potential is slippery ground, because it might easily slide into a mixed counterfactual account of the kind discussed earlier. It would be easy to make this slide because in any case where a person would have been better off in the absence of a specified causal process, it is presumably possible to say that there has been interference with some kind of "potential." It would be implausible to claim that all such interferences are harm. However, once we introduce the idea of a potential it might be difficult to avoid biting that particular bullet, and once the bullet has been bitten it would be implausible to claim that the harmful character of every such interference depends on an historical rather than a counterfactual inquiry. In order to characterize interference with a natural potential as an historical worsening, I think that much must therefore rest on the idea that the potential is a *natural* one, meaning one that inheres in the nature of the being affected.²⁷ I am not sure at the moment how to develop this idea further; as I said earlier, this general approach to the problem cases is offered only tentatively.

26. Cases of this kind were pointed out to me independently by Richard Arneson and Tony Honoré.

27. As was noted in Part II, Feinberg believes that a doctor who improves the condition of a patient, but not as much as she should have done, has harmed the patient. In my opinion, this is not a case of harm but rather a failure to confer a benefit. It does not matter, for purposes of this analysis, that the doctor was under a duty to confer the benefit. I am assuming that the doctor did not impede a natural potential to recover, but simply failed to make things better in a situation where they would not have gotten better on their own. See further *supra* note 25.

III. SHIFFRIN'S CRITIQUE OF THE COMPARATIVE MODEL

Both Feinberg's counterfactual analysis of harm and the historical worsening account depend on a comparison between two states of affairs. The former compares what has happened with what would have happened, while the latter compares what has happened with the status quo ante. Seana Shiffrin takes both types of account to be particular versions of what she calls the "comparative" model of harm, and she offers a critique of that model which she maintains applies to both versions. She takes Feinberg's analysis of harm as her primary target, but insofar as the problems she discerns with the comparative model flow from comparison *tout court*, rather than from counterfactual comparison in particular, the critique is *prima facie* applicable to the historical worsening account as well. Shiffrin describes the comparative model in the following terms:

[M]any regard harms and benefits as though they represent two ends of a scale, like the scale of positive and negative numbers. Benefits are thought to be just like harms, except that harms are bad and benefits are good. On Feinberg's natural and attractive interpretation of this symmetrical picture, harms involve the setback of one's interests, whereas benefits involve the advancement of one's interests along a sliding scale of promotion and decline. To evaluate whether an event has benefited or harmed a person, one compares, with respect to the fulfillment of his interests, either his beginning and his end points (historical models), or his end point and where he would have been otherwise (counterfactual models). If he has ascended the scale (either relative to his beginning point or alternative position), then he has been benefitted [sic]. If he moves down, then he has been harmed. Either way, one arrives at an all-things-considered judgment that either harm or benefit (but not both) has been bestowed. Thus, because he has been overall benefited, he has not been harmed.²⁸

Shiffrin argues that there are many problems with the comparative model, thus understood. First, it fails to explain certain "deep asymmetries" between benefit and harm. Shiffrin claims that we often look upon failing to be benefited as morally and substantively less serious than both being harmed and being saved from harm. She notes that this asymmetry is impossible to make sense of on a counterfactual model, but that even on an historical model the distance between the start and end points may be too small to explain the asymmetry properly. Second, either the historical or the counterfactual version of a comparative account will have to say that persons who have been moved to status *x* from different ends of the scale have been benefited or harmed respectively, regardless of their

28. Shiffrin, *supra* note 3, at 121 (footnote omitted).

starting points. Suppose *A* is (or would otherwise have been) at $x - 2$, and is moved to x . *B*, on the other hand, is (or would otherwise have been) at $x + 2$, and is also moved to x . The comparative model is committed to saying that *A* has been benefited and *B* harmed, even though they are now identically situated. Because of what she takes to be the asymmetry of benefits and harms, Shiffrin regards this as a problematic outcome. It would be even more problematic, she believes, if *A* is moved from $x - 4$ to $x - 3$, whereas *B* is moved from $x + 2$ to $x + 1$. In this hypothetical, the comparative model tells us that *A* is benefited and *B* harmed, even though *B* is better off, all things considered, than *A*. “If this were so, why should harm, per se, in this sense, be a special subject of moral concern and have greater priority than failures to be benefited?”²⁹

Shiffrin offers a number of other criticisms of the comparative model of harm, but all of them are related to her central thesis that that model cannot explain the asymmetry that she says exists between harms and benefits. She therefore offers a rival account of harm, the most significant aspect of which is its noncomparative character:

Typically, harm involves the imposition of a state or condition that directly or indirectly obstructs, prevents, frustrates, or undoes an agent’s cognizant interaction with her circumstances and her efforts to fashion a life within them that is distinctively and authentically hers To be harmed primarily involves the imposition of conditions from which the person undergoing them is reasonably alienated or which are strongly at odds with the conditions she would rationally will On this view, pain counts as a harm because it exerts an insistent, intrusive, and unpleasant presence on one’s consciousness that one must just undergo and endure. Disabilities, injured limbs, and illnesses also qualify as harms. They forcibly impose experiential conditions that are affirmatively contrary to one’s will; also, they impede significantly one’s capacities for active agency Death, too, unless rationally willed, seriously interferes with the exercise of agency.³⁰

I should note, to begin, that I have some doubts about Shiffrin’s introduction into her account of the idea that harm consists of conditions that are, inter alia, ones from which one is alienated or which one did not or would not rationally will. This strikes me as problematic for the reason that it makes harm potentially too subjective a notion. If, for whatever reason, one is not alienated from a condition that would otherwise count as harm, or if one rationally willed the condition (say, via consent), should we therefore conclude that the person is not suffering harm? My inclination is to say no; while one can consent to

29. *Id.* at 122. Shiffrin notes that a loss as such can be a morally significant harm if there were an expectation or personal investment involved, but argues, correctly, that the comparative model does not make anything turn on the fact of expectation or investment. *Id.*

30. *Id.* at 123–24.

harm, such consent does not (except in special cases)³¹ transform the harmful condition into a nonharmful one. I will say no more about this issue here, however. My primary concern is not with Shiffrin's alternative account of harm as such, but rather with its relevance to her critique of the comparative model.

In the displayed passage quoted above, Shiffrin seems to use the term "imposed" in two different ways. Early in the passage she writes that "[t]o be harmed primarily involves the imposition of conditions from which the person undergoing them is reasonably alienated." This suggests, to my mind, that what is being imposed are such conditions as a disability, an injured limb, or an illness. Because these conditions would presumably be "imposed" by means of a causal mechanism, this understanding of "imposition" is completely consistent with the historical worsening condition. Later in the passage, however, Shiffrin speaks of disabilities, injured limbs, and illnesses as *themselves* imposing experiential conditions that are affirmatively contrary to one's will. This suggests that, on Shiffrin's view, disabilities, injured limbs, and other harms are *simply conditions*, meaning states of being that involve or lead to (impose) further, experiential conditions such as unpleasant mental states or a subnormal capacity for agency. To count as harm, these states of being need not have come about as the result of any historical worsening. On this view, a disability or an illness with which one was born is just as much a harm as a disability or an illness that befell one during one's lifetime.

I am not sure which of these two senses of "imposition" Shiffrin would accept as giving the appropriate content to her affirmative account of harm. I believe, however, that the correct approach would be to incorporate the historical worsening condition into the account, thereby treating harms as states or events that are "imposed" in the sense that they are historical worsenings that were caused by prior states or events. Shiffrin seems to assume that the historical version of the comparative model must treat any worsening of an interest as a harm (and any improvement as a benefit). This does seem to be Feinberg's position (at least in the sense that the process of "netting out" takes account of all setbacks to interests on the minus side, and of all advancements on the

31. The kind of special case I have in mind would be the harm that Duff argues is associated with rape; unlike most harms, this harm is integrally bound up with the human action that created it. See DUFF, *CRIMINAL LIABILITY*, *supra* note 22, at 111–12. But if the sexual act is in fact consented to, then the harm in question does not exist at all.

plus side). But comparative models are not, in general, committed to that idea. The historical worsening account need only say that a worsening is a necessary condition of harm, not a sufficient condition. It is perfectly consistent with such an account that harms, to be harms, must also meet other conditions, such as, for example, those set out in Shiffrin's own affirmative analysis (that the worsened state be such that one is alienated from it or be unable rationally to will it).

Shiffrin argues that "we often consider failing to be benefited as morally and significantly less serious than both being harmed *and* not being saved from harm."³² She notes, correctly, that counterfactual comparison models of harm cannot distinguish at all between harming and failing to benefit. She further notes, again correctly, that historical comparison models do have the resources to draw this distinction, but goes on to argue that "the distance between the end points that make it a harm rather than a failure to be benefited may be rather too small to account for the strength of our asymmetrical reactions."³³ The point here would seem to be that the setbacks to interests that we regard as harms are often more serious than failures to confer a benefit because harms, unlike the general category of failing to benefit, necessarily involve particularly serious evils, such as illness or personal injury. I do not think that this conceptual claim is correct, but even if it is, we could, as was noted in the preceding paragraph, respond to the concern by supplementing the historical worsening condition rather than by rejecting it; we could insist that there are other necessary conditions that must be met before harm can be said to occur. Again, one possibility is Shiffrin's own affirmative analysis, although for the reasons stated earlier I have some doubts about the viability of that approach. It should also be noted that while we have or should have the same level of moral concern about serious evils like illness even when they do not come about as a result of an historical worsening (for example, when they involve congenital conditions), it does not follow that such evils simply are harms in the absence of an historical worsening.

As I have already noted, the historical worsening model of harm is not committed to the view that all historical worsenings are harms. Even while conceding that point, though, I think Shiffrin's case against the comparison model in general (and hence the historical worsening model in particular) is overstated. Consider her argument about two persons, *A* and *B*, who are both moved to interest level x , where the former was previously at level $x - 2$, while the latter was at level $x + 2$. The comparative model is committed, Shiffrin says, to saying that *A* has been

32. Shiffrin, *supra* note 3, at 121.

33. *Id.* at 122.

benefited and *B* harmed, even though the two are now identically situated. But what is so obviously problematic about saying that? Suppose that *A* was suffering from a severe form of schizophrenia, while *B* was suffering from a mild form. Someone has done something to *A* to alleviate her condition somewhat, while someone else has done something to *B* to worsen his, so that they are now suffering from the disease to more or less the same degree. Is it not correct to say that *B* was harmed while *A* was benefited? Shiffrin asks, “[W]hy should harm, per se, in this sense [that is, in the comparative sense], be a special subject of moral concern and have greater priority than failures to be benefited?”³⁴ It is not entirely clear to me, however, that harm *does* have greater priority just because it is harm. What has priority, in the sense I think Shiffrin has in mind, is the existence of certain evils, which may or may not be associated with harm. It is also worth noting that there is, of course, another source of moral concern in the worsening of *B*'s condition, and that is the fact that it flowed from an act rather than an omission (or, perhaps, from a “doing” rather than an “allowing”). I believe that distinctions of this kind are doing far more work in our intuitions than Shiffrin acknowledges.

Whether or not we accept Shiffrin's particular account of harm as involving a condition from which one is alienated or which one could not rationally will, we still must ask whether or not a philosophical account of harm should limit the historical worsenings that count as harm to those that involve such discrete and serious evils as disease, disability, personal injury, or death. It seems to me that a satisfactory account should not be so limited. The ordinary concept of harm is properly applied, for example, to the model who was prevented from entering the Miss America pageant even though she was not subjected to an evil of the kind just mentioned. There may well be a *de minimis* constraint on the extent or seriousness of the setback to an interest that must obtain before the setback will count as harm, but that is a different matter. Even beyond that question, however, there remains more to be said about which interferences with which interests should count as harm, and it is to that general topic that I now turn.

34. *Id.*

IV. HARM, INTERESTS, AND RIGHTS

While I have rejected Feinberg's counterfactual analysis of harm, I have accepted his more fundamental point that a harm is a setback to an interest. I have further argued for the thesis that such a setback should be understood as an historical worsening. This invites the question: Which are the interests that, when set back, give rise to harm? Is any interference with any interest properly designated a harm?

Let me begin by considering a hypothetical discussed by Feinberg. Suppose a rescuer rescues a seriously endangered person, but finds that in order to effect the rescue and save the imperiled person's life, he necessarily has to break the other's arm in the process. One question that arises is this: Is the rescuer liable for the broken arm if the rescuee chooses to sue? But the prior, more fundamental question is the following: Did the rescuer harm the rescuee? In answer, Feinberg offers this argument:

[T]he broken-armed plaintiff suffered a harmful condition with respect to his arm, but the rescuer-defendant did not cause a condition that was harmful on balance, offset as it was by the overriding benefit of rescue, and he cannot be said, therefore, to have *harmed* the plaintiff (in the relevant full sense) at all.³⁵

Shiffrin discusses this case and suggests, in light of her own affirmative discussion of harm, that we should not deny that the necessarily broken limb is a form of harm: "It imposes a condition of disability and inflicts pain. It seems to meet the criteria of harm, then, and does so irrespective of the concomitant benefits delivered alongside it; on these criteria, it can be harm even if, in some overall sense, the event makes the person better off."³⁶

Even if we do not accept Shiffrin's particular analysis of harm, she nonetheless seems exactly right in her analysis of the limb-snapping rescuer. Feinberg's claim to the contrary notwithstanding, there does not seem to be some overall interest into which all harms and benefits can be factored, making possible a net judgment about the harm (or benefit) that any given person has received overall.³⁷ The general interests that might

35. FEINBERG, *supra* note 2, at 27.

36. Shiffrin, *supra* note 3, at 125.

37. In his contribution to this symposium, Leo Katz offers a very interesting discussion of this kind of case, as well as of the general issue of netting out benefits against harms. See Leo Katz, *What to Compensate? Some Surprisingly Unappreciated Reasons Why the Problem Is So Hard*, 40 SAN DIEGO L. REV. 1345 (2003). I cannot discuss Katz's approach here, except to make the very general observation that I do not think he distinguishes sufficiently clearly between the question of netting out, which concerns the dual issues of the occurrence of harm and the quantification of damages, from the question of whether or not a harmful act was justified by the benefits it also produced.

plausibly be thought to be candidates for this role are, presumably, welfare and autonomy. While autonomy strikes me as the more likely possibility of the two, in the end I do not think our ordinary concept of harm permits this kind of reductionism with respect to either of these general interests. This is, in part, a point about the concepts of harm and benefit themselves. While questions about the individuation of harms can and do arise, I believe that our *general* tendency is to treat setbacks to interests even in the same general category, for example, physical injuries to different parts of the body, as distinct harms. This is true even though personal injuries all tend to set back the same set of interests, namely, autonomy and freedom of movement, health, life, the physical integrity of the body, the interest in not experiencing pain, and the interest in not experiencing mental or emotional distress. Obviously, particular physical injuries set back various of these interests in different degrees, and some injuries will leave some of these interests unaffected. Perhaps that is part of the reason that we tend not to lump even concurrent instances of physical injury, let alone setbacks to interests that are less obviously systematically related, into a single aggregate harm. A fortiori, we do not net out harms against benefits. The most important reason not to lump or net out, however, is surely substantive in character. We have no good reason to think that the myriad array of interests that are subject to harm and benefit—those just mentioned in connection with physical injury, for a start, and many others as well, such as dignity, privacy, the interest that I have in living according to proper values, and interests of various kinds that I have in tangible and intangible property—can all be reduced to a single underlying interest of any kind. (This is not to deny, of course, that there is much overlap among the relevant interests.) As Shiffrin notes, the better way to view the limb-snapping case is that, in the particular circumstances, the bestowal of the benefit (saving the rescuee's life) might *justify* the breaking of his arm for purposes of criminal or civil liability.³⁸

I suggested in the preceding Part that all harms are historical setbacks to interests, but further noted that there was no reason to think that all historical setbacks to interests are harms. I would like at this point to

38. Shiffrin, *supra* note 3, at 126. Having discussed the netting-out interpretation of Feinberg's explanation of the limb-splitting example, Shiffrin goes on to interpret him, more charitably, as putting forward the idea of justification of harm instead. This is a very plausible way to understand the particular example, but it seems clear that what Feinberg himself had in mind was netting out, not justification.

engage in a speculative and somewhat tentative inquiry that takes this thought as its starting point. It seems reasonable to think that there is some set of core or primary interests, all of which are roughly concerned with our well-being and how our lives go, that are, so to speak, the main targets of harm. These core interests may overlap to some extent, and some may be partially derivable from others. For present purposes, there is no need to try to ascertain exactly what the core interests are, but it seems plausible to think that many of the interests mentioned in the preceding paragraph will figure on any acceptable list. Assuming there is such a list of core interests, can we conclude that those are all the interests that we have? I think the answer is clearly no. Consider, for example, my interest in not being physically injured. As emerged in earlier discussion, this interest probably represents an amalgam of several different interests, but nothing will be lost for present purposes by treating it as discrete and stand-alone in character. If I have an interest in you not physically injuring me, surely I also have an interest in you not *trying* to physically injure me, as well as an interest in you not engaging in actions that would *risk* you physically injuring me. Does this mean that if you try to injure me but fail, or if you subject me to a risk of a physical injury that does not materialize, then I have been harmed? It is of course possible that the trying or risking will interfere with one or another of my core interests, say by causing me severe fright or subjecting me to indignity. In order, therefore, to focus on the precise question at issue, let me assume that no such core interest is set back by the trying or risking. The question remains: Was I harmed simply because you acted contrary to the interest that I have in you not attempting to physically injure me (or the interest that I have in you not subjecting me to certain kinds of risk)?

I have argued elsewhere³⁹ that because of the peculiar epistemic character of risk, subjecting another person to risk cannot, in and of itself, constitute a harm. I would now want to qualify that conclusion by saying that risking cannot be regarded as adversely affecting any interest that has a strong or plausible claim to be in the set of core or primary interests. The further suggestion I am advancing is that the very existence of a set of core interests will give rise to secondary interests which will at least sometimes be second-order interests, meaning interests that are defined recursively. The interest I have that you not subject me to certain kinds of risk of physical injury is just such a second-order interest. Do you harm me by acting contrary to that interest? My strong intuition is that you do not. The idea of a core of protected interests is, I

39. Stephen R. Perry, *Risk, Harm, and Responsibility*, in *PHILOSOPHICAL FOUNDATIONS OF TORT LAW* 321 (David G. Owen ed., 1995).

believe, part of our concept of harm, although we may disagree about which interests lie inside the core and which outside. (Consider moral or aesthetic offence, for example.) Although I think Shiffrin was mistaken to ignore the historical worsening aspect of harm and thereby to equate harm with certain evil conditions alone, the fact that she focused on *certain* evil conditions I take to be itself a recognition of this truth. Once we have acknowledged, as I think the concept of harm itself forces us to do, that some adverse effects on interests are harms and some are not, the distinction between core interests on the one hand, and secondary interests on the other, becomes a very natural way to further systematize and make sense of this feature of the concept.

It is worth noting two further points about the suggested distinction between first- and second-order interests. The first point is that, theoretically, there could be higher-order interests still, and the more rarefied these become, the less likely we are to say that interference with them constitutes harm. This offers at least some support to the idea that *one* cutoff between interests that can be harmed and those that cannot—obviously, I do not need to claim that this is the only such cutoff—should be drawn along the lines of core interests on the one hand, and higher-order interests on the other. The second point is that, even if it is true that interference with higher-order interests does not generally constitute harm, it does not follow that higher-order interests (or secondary interests generally) cannot be subject to a right. I assume, for purposes of the present discussion, an interest theory of rights along the lines proposed by Joseph Raz.⁴⁰ There is nothing odd about the idea, and indeed it seems intuitively correct, that I have a right that you not try to physically injure me, even though a violation of that right which did not cause me physical injury would not itself be harm. The right is based on my interest that you not try to physically injure me, which is itself a second-order interest recursively derived from my first-order, core interest in not being physically injured. I can have a right that you not behave in a certain way that is not, in and of itself, harmful to me, because your behaving in that way might cause me to be harmed.

Obviously a great deal more remains to be said about the relationship among harm, interests, and rights, and I cannot begin to do justice to the topic here. But I would like to add a few remarks that at least suggest

40. RAZ, *supra* note 19, at 165–92. I have benefited a great deal from Raz's discussion of the relationship between rights and interests.

the direction that further discussion should take.⁴¹ Why, it might be asked, should a secondary interest ever be subject to a right, if interference with that interest is not itself harm? The answer is that protecting secondary interests often has instrumental importance. This is true, for example, of the second-order interests that I have that others not try to physically injure me or subject me to the risk of physical injury. Protecting these interests will generally, although by no means invariably, help to ensure that my core interest in not being physically injured does not suffer a setback. On an interest theory of rights, *A* has a right against *B* only if, inter alia, some interest of *A*'s is a sufficient reason to hold *B* to be under a duty.⁴² Clearly, the ultimate justification of a right that others not try to injure me or subject me to a risk of injury is my core interest in not being physically injured. But recognizing that the justification of a right sometimes runs through a secondary, instrumental interest shows that it is conceptually possible that a right can be violated, and an interest set back, without any harm occurring.⁴³ Once this point is recognized, it becomes clear that one cannot, for example, simply help oneself to the conclusion that risk is harm by mere virtue of the fact that risking sets back an interest. All too often, I believe, the justification for this conclusion takes essentially this form,⁴⁴ when what is required is a substantive moral argument to the effect that a secondary, instrumental interest is, for some purposes at least, morally on a par with a fundamental interest. In the absence of such an argument, there is simply no reason to think that risking is harming.

Something similar holds for affirmative rights. Consider the important

41. These remarks were prompted by the comments of John Goldberg and others on the version of this Article that I presented at the symposium on compensation and the Oxford Jurisprudence Colloquium.

42. RAZ, *supra* note 19, at 166. I should note that the distinction between core and secondary interests is related to, but not the same as, the distinction that Raz draws between core and derivative rights. *See id.* at 168–70.

43. Often, of course, and perhaps usually, a right is justified directly by the core interest. When this is the case, a violation of the right is itself harm. If *A* violates a right that *B* has not to be injured, then *A* has harmed *B*. But there are important examples, in addition to trying and risking, when this is not the case. Consider, for example, the interesting case of trespass, which is raised by John Goldberg in his comment. *See* Goldberg, *supra* note 19, at 1321–22. The right that others not trespass on one's land protects an interest in the exclusive possession of the land. It seems to me that this interest is secondary and instrumental; we only protect it in order to protect other interests, such as autonomy and privacy interests. Goldberg is thus correct to conclude that there can be harmless trespasses. On the general issue of harmless wrongs, see also Heidi M. Hurd, *What in the World Is Wrong?*, 5 J. CONTEMP. LEGAL ISSUES 157, 212–15 (1994).

44. *See, e.g.,* Wright, *supra* note 24, at 1814–16. An exception to this generalization is Matthew Adler's very interesting argument in Matthew D. Adler, *Risk, Death and Harm: The Normative Foundations of Risk Regulation*, 87 MINN. L. REV. 1293 (2003). Unfortunately, I cannot discuss Adler's approach to these issues here.

cases of promising and contracting. Suppose that *A* promises *B* that, on a certain date and at a certain time, *A* will turn a pirouette. So far as the content of the promise is concerned, we can assume that nothing turns on whether or not *B* knows that *A* has turned the pirouette. But even if *B* never finds out whether or not *A* has done the promised act, *A* nonetheless has an obligation to turn a pirouette, and *B* has a corresponding right that *A* perform this action. It is possible, of course, that *A*'s failure to turn a pirouette will harm *B*. *B* might rely in some way on *A*'s turning the pirouette, or *B* might suffer disappointment when he discovers that *A* did not do as she promised. But let us assume that *B* did not so rely, and he never finds out whether or not *A* performed as promised. *B* nonetheless has an interest that *A* perform a pirouette at the specified time and place. As Raz has argued, every person has an interest that promises made to him will be kept, and *B*'s interest in *A*'s performing a pirouette is just a special instance of this more general interest.⁴⁵ In contract law, both the general and the special interest are known, in a harmless ambiguity, as the expectation interest. This is the interest that is protected by *B*'s right to performance.⁴⁶ For obvious practical reasons, promises do not usually get made unless performance of the promise is likely to benefit the promisee.⁴⁷ By "benefit," I mean the advancement of one or more of the promisee's core interests. But the expectation interest itself is best viewed as a secondary, instrumental interest. Thus, advancement of the expectation interest does not necessarily confer a benefit in the sense of advancing a core interest, and it is plausible to think that this is true in the pirouette case. Still less does a setback to the expectation interest, through a failure to fulfill one's duty to perform as promised, in itself constitute harm. Thus, it would be very implausible to think that if *A* violates *B*'s right that *A* turn a pirouette, *A* has ipso facto harmed *B*.

V. HARM AND COMPENSATION

I come, finally, to the relationship between harm and compensation, particularly in the context of the law of torts. I have argued in preceding Parts that a necessary condition of harm, and indeed one of its principal features, is that there must have been an historical setback to one of the

45. Cf. RAZ, *supra* note 19, at 175. The alternative view, which Raz rejects (I believe correctly), is that the right is based on the promisee's interest in the promised act.

46. See generally Peter Benson, *Contract*, in *A COMPANION TO PHILOSOPHY OF LAW AND LEGAL THEORY* 24 (Dennis Patterson ed., 1996).

47. Cf. RAZ, *supra* note 19, at 175.

relevant interests of the person who allegedly suffered harm. Assume for the present that this understanding of harm is correct. If *A* wrongfully harms *B*, and if harm is properly understood as an historical worsening, why should corrective justice not require full compensation for that very harm? Consider a variation on the *Kingston* case that involves preemptive causation. *A* wrongfully sets a fire which burns down *B*'s house. If *A*'s fire had not had this effect, then a naturally caused fire (say, one set by lightning) would have burned the house down a few moments later. The courts have generally said in this situation that *B* is not entitled to compensation, or if she is, it is only for the few moments of house-possession of which *A* deprived her.⁴⁸ The general position was set down in 1880 by Lord Blackburn in the *Rawyards* case:

I do not think there is any difference of opinion as to its being a general rule that, where any injury is to be compensated by damages, in settling the sum of money to be given for reparation of damages you should as nearly as possible get at that sum of money which will put the party who has been injured, or who has suffered, 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.⁴⁹

Determining compensation, in other words, requires a counterfactual inquiry. The idea is to put the plaintiff in the position that he or she would have been in had the tort not occurred, to the extent that this can be done with money.

Let me return to my question about the variant on *Kingston* just described. If *A* burned down *B*'s house, why should he not pay for it, even though it would have burned down anyway? Shouldn't corrective justice require him to correct the harm he caused, regardless of what else might have happened if he had not committed his wrong? The answer cannot be that the *Rawyards* principle is a conceptual implication of the concept of compensation, because that is clearly not true. A principle that looked to the status quo ante as a baseline of compensation rather

48. Cf. *Dillon v. Twin State Gas & Elec. Co.*, 163 A. 111, 114–15 (N.H. 1932).

49. *Livingstone v. Rawyards Coal Co.*, [1880] 5 A.C. 25, 39 (H.L.). John Goldberg has suggested that the language in *Rawyards* might be understood as indirectly invoking an historical worsening approach to quantifying damages and notes further that many jury instructions refer explicitly to the idea of restoring the plaintiff to the status quo ante. See Goldberg, *supra* note 19, at 1321 & n.19. I believe that a detailed examination of subsequent English cases would bear out the claim that English law is to be understood as requiring damages to be assessed counterfactually rather than historically, but I acknowledge that in American law there is some disagreement, and perhaps equivocation, about this issue. For a statement of an American equivalent of the *Rawyards* principle, see *McDougald v. Garber*, 536 N.E.2d 372, 374 (N.Y. 1989) (“The goal [of an award of compensatory damages] is to restore the injured party, to the extent possible, to the position that would have been occupied had the wrong not occurred.”) (citation omitted).

than to what would have happened had the wrong not occurred is perfectly coherent and, indeed, has much to be said for it in a substantive sense. The answer also cannot be that to give *B* compensation for her house would be to give her a windfall, because in the absence of a justificatory argument for one result or the other, it is just as much a windfall for *A* not to have to pay compensation.

The *Rawyards* principle does seem to me to state a generally correct understanding of how damages should be awarded in torts, or rather it states a first approximation of such an understanding, for there are exceptions. What is the basis for saying this? I am by no means satisfied with my answer, and I wish I had a better one. The only justification that I can offer takes roughly the following form. An award of compensatory damages is, in the first instance, supposed to make good the historical harm that one party has caused another. But such an award is obviously not intended literally to roll back the past. Given that impossibility, there are a number of factors that are appropriately taken into account in quantifying damages in monetary terms. One of these is what we might call “net-worseoffness,” which is the difference, as ascertained by a counterfactual inquiry, between the plaintiff’s current circumstances and the circumstances he or she would have been in had the wrong not occurred. Given that *B*’s house would have burned down anyway, so that even though *A* has caused *B* harm he has not actually left her worse off, it is a reasonable view of what justice requires between the parties to say that *A* does not have to pay for the house. Discounting for contingencies would be justified along similar lines. I emphasize that this is a substantive and contestable moral judgment, not a conceptual truth of any kind. It is also important to see that while net-worseoffness is a morally relevant factor in the quantification of damages, it is only one among others. Thus the *Rawyards* principle does not state a hard and fast rule. Out of an abundance of caution I should also emphasize that net-worseoffness is not, for the reasons that were given in Parts II and IV, an issue that arises in the determination of whether or not there has been harm. It is, rather, a factor to be taken into account in the assessment of damages.

To see why net-worseoffness is only one factor among others in the quantification of damages, consider the basic fact situation of *Kingston* again. Two fires joined and burned down the plaintiff’s house, where either fire alone would have been sufficient to cause that outcome. The defendant set one of the fires, whereas the other was of unknown origin.

The court said that if the other fire could be shown to be of natural origin, *A* would not have to pay *B* compensation. This is a clear application of the *Rawyards* principle. But the court then went on to say that if, as it assumed was in fact the case,⁵⁰ the second fire had been tortiously set, then both tortfeasors could be held jointly and severally liable for the harm. Given that only the one tortfeasor was before the court, that person could be held fully liable. The court refused, in other words, to apply the *Rawyards* principle to this set of facts. The injustice of allowing a tortfeasor to escape liability by pointing to what another tortfeasor had done was held to be of sufficient moral weight to overcome the prima facie assumption that *A* does not have to pay compensation for a kind of harm that would have occurred anyway. It is very arguable that the same result should be reached when the second tortfeasor did not causally contribute to the actual harm, but would have caused similar harm had the first tortfeasor not already done so. Because the second tortfeasor did not actually cause any harm, he cannot be held liable. But it can nonetheless be argued that the first tortfeasor should have to pay full compensation because, one way or another, the harm the plaintiff was bound to suffer would have been tortiously caused, and it would be unjust to let her go uncompensated. And this is the result that at least some courts have seen fit to reach.⁵¹

What, then, are we to say of the venerable doomed airplane? Most commentators assume that the taxi driver does have to pay compensation, and my intuitions run in the same direction. It should be emphasized that the case does not raise the *Rawyards* principle at all; there is no question of even trying to put the plaintiff in the position he would have been in had the tort not occurred (namely dead). Rather the case raises the problem of offset: When, and under what circumstances, should the benefits that flow from the defendant's tortious actions serve to mitigate damages? It should also be emphasized that offset, being a question of damages, is different from the question of whether the occurrence of harm depends on a netting-out effect, and different again from the question of whether or not an accompanying benefit *justifies* a harmful action, that is, renders it nonwrongful.⁵² The most plausible understanding of the limb-splitting rescuer, for example, is that his action was justified by the fact that causing the harm avoided an even greater harm, namely

50. The court reversed the burden of proof on this issue. It is not obvious that it was correct to do so, but that point is not directly relevant to the question currently under consideration.

51. See, e.g., *Baker v. Willoughby* [1970] A.C. 467, 494 (H.L. 1969).

52. Shiffrin offers a very interesting discussion of the justification question. See Shiffrin, *supra* note 3, at 125–31.

death. The question of damages would thus not even arise.⁵³ If, in the doomed airplane hypothetical, the taxi driver had intentionally injured the would-be passenger as the only available means of preventing him from getting on the airplane, the issue of justification would have to be addressed.⁵⁴ But that issue clearly does not arise when the driver negligently injures the would-be passenger with no inkling that the plane is doomed to crash.

The benefit in the doomed airplane hypothetical is, of course, the avoidance of the harm of death. The *Restatement* says there should be offset “to the extent that this is equitable.”⁵⁵ Although not very helpful in concrete terms, this nonetheless strikes me as an appropriate way to put the point: Offset is another factor that is morally relevant to the quantification of damages, but it is one that requires substantive moral judgments that may be quite controversial and that may shift from case to case. I do not have a knockdown argument one way or another in the case of the doomed airplane, but I believe the equities run against mitigation because the benefit is too much of a coincidence, and too little connected to the defendant’s actual actions, to be appropriately treated as an offset. The case is reminiscent in this respect of the coincidence cases that arise under the rubric of proximate cause.⁵⁶ As a result of the defendant’s negligence, the plaintiff is, say, placed under a tree that just happens to fall at the precise moment that he or she is there. The coincidence rules out liability in such a case, and my sense is that it should similarly rule out offset in the case of the doomed airplane.⁵⁷ But I concede that one could probably make a respectable argument that would come out the other way.

53. This is not to say that there can never be liability for a justified act. See *Vincent v. Lake Erie Transp. Co.*, 124 N.W. 221, 222 (Minn. 1910).

54. Leo Katz discusses this variation of the doomed airplane hypothetical. See Katz, *supra* note 37, at 1348.

55. RESTATEMENT (SECOND) OF TORTS § 920 (1979).

56. See, e.g., *Berry v. Sugar Notch Borough*, 43 A. 240 (Pa. 1899).

57. At the conference on compensation, some participants dubbed this the “symmetry” principle.

