1975

# Concerning Cause and the Law of Torts: An Essay for Harry Kalven, Jr.

Guido Calabresi
*Yale Law School*

# Concerning Cause and the Law of Torts: An Essay for Harry Kalven, Jr.*

*Guido Calabresi*†

*In 1960 I walked into an office at The University of Chicago Law School. There I found Walter Blum and Harry Kalven. They had just read a draft of what was to become my first article. Harry greeted me with:* "it's all wrong . . . *but I wish* I *had written an article like that when I was your age!" This began the debate. Blum and Kalven delivered the Shulman Lectures at Yale,* Public Law Perspectives on a Private Law Problem—Auto Compensation Plans*; I struck back in* Fault, Accidents and the Wonderful World of Blum and Kalven*; but they had the last laugh in* The Empty Cabinet of Doctor Calabresi. *I believe that the law of torts benefited from that hard fought polemic. I know that I, as a young scholar, could not have had a tougher, or more loving, initiation to scholarship. Fifteen years have passed and Harry is no more. Because I think that he would disagree with this paper as much as he did with my first, I gratefully dedicate it to his memory.***

Sooner or later most commentators in the field of torts find it necessary to deal explicitly with the problem of causation. Indeed, so much has been written about the significance of a causal requirement in torts, on both a philosophical and a hornbook level, that another essay may seem superfluous. At this point, a digest, criticism, or even analysis of those previous writings would not be especially useful.[1] Moreover, the philosophical significance of the con-

---

* This article is derived from the Harris Lectures, given at the University of Indiana School of Law (Bloomington) in October 1974. The lectures will be published in full, in book form.

† Professor of Law, Yale University.

** W. BLUM & H. KALVEN, PUBLIC LAW PERSPECTIVES ON A PRIVATE LAW PROBLEM—AUTO COMPENSATION PLANS (1965); Blum & Kalven, *Public Law Perspectives on a Private Law Problem—Auto Compensation Plans*, 31 U. CHI. L. REV. 641 (1964); Calabresi, *Fault, Accidents, and the Wonderful World of Blum and Kalven*, 75 YALE L.J. 216 (1965); Blum & Kalven, *The Empty Cabinet of Dr. Calabresi: Auto Accidents and General Deterrence*, 34 U. CHI. L. REV. 239 (1967).

[1] The influence of many of these writings on my thougbt should be readily apparent. This paper is not the place to attempt a bibliography of even the major works on causation and

cept of causation is something that I am neither inclined nor trained to discuss. Yet, for all of that, one facet of the relationship of causation to tort law does deserve exploration.

In recent years, tort law has been reexamined with certain ends or functions consciously in mind. Commentators have asked how the law of torts determines what injuries are worth avoiding, how it controls what categories of people bear the burden of those injuries that do occur and the related burden of avoiding those injuries deemed worth avoiding, and how it serves to encourage or require the spreading of such burdens. Finally, commentators have considered how these goals relate to the concept and language of justice—the test, however vague and uncertain, by which any area of law must be judged.[2]

This functional approach has come to dominate American tort scholarship. Even its severe critics begin with this framework in order to point out the limitations they perceive in it. Yet those, myself included, who have been the most assiduous in furthering this approach have not explicitly considered the role causation should play when tort law is examined in terms of deterrence, spreading, and distributional goals. What significance does the requirement of causation have for those ends? Can causation serve such goals, or is it, as some have regarded the fault principle, an impediment to their achievement? Would an explicit analysis of causation in the light of such goals help explain those cases that have seemed paradoxical under traditional analysis? These questions, along with the closely related question of what role causal language should play in a tort system eschewing the fault principle, deserve consideration.

All of these questions cannot be explored in this short paper.[3] My more modest task here will be to analyze causation in terms of

---

torts. I would, however, single out for particular mention Leon Green, H.L.A. Hart and A.M. Honoré, Fleming James, Robert Keeton, and Wex Malone, because their works were especially significant in the development of my ideas on the subject.

[2] *See, e.g.*, W. Blum & H. Kalven, Public Law Perspectives on a Private Law Problem (1965); G. Calabresi, The Costs of Accidents: A Legal and Economic Analysis (1970); Calabresi, *Optimal Deterrence and Accidents*, 84 Yale L.J. 656 (1975); Calabresi & Melamed, *Property Rules, Liability and Inalienability*, 85 Harv. L. Rev. 1089 (1972); Posner, *Strict Liability*, 2 J. Legal Studies 205 (1973); Posner, *A Theory of Negligence*, 1 J. Legal Studies 29 (1972); *cf.* Epstein, *Intentional Harms*, 4 J. Legal Studies 391 (1975); Epstein, *Defenses and Subsequent Pleas in a System of Strict Liability*, 3 J. Legal Studies 165 (1974); Epstein, *A Theory of Strict Liability*, 2 J. Legal Studies 151 (1973); Epstein, *Pleadings and Presumptions*, 40 U. Chi. L. Rev. 556 (1973); Fletcher, *Fairness and Utility in Tort Theory*, 85 Harv. L. Rev. 537 (1972).

[3] These questions will, however, be considered at length in the book of which this article is only a part.

deterrence, spreading, and distributional goals and then to apply my observations to a few classic causation cases. Throughout this enterprise only the functions of causal language will be considered; deeper questions about the meaning of cause cannot and will not be discussed. Accordingly, the definitions of cause used in this paper will not even begin to cover the field. Instead, they will be pedestrian definitions designed to do no more than further analysis of the question I have posed: what role does causal language play in the achievement of certain goals that have come to be accepted as crucial to the law of torts?

To accomplish this function-oriented task it is useful to distinguish three concepts of "cause": "causal link," "*but for* cause," and "proximate cause."

The first concept, as I shall use it, is entirely predictive and empirical. There is a causal link between an act or activity and an injury when we conclude on the basis of the available evidence that the recurrence of that act or activity will increase the chances that the injury will also occur. There is a causal link between the failure to keep a light on at the top of a flight of stairs and a person falling down those stairs, whether or not a particular fall was in some other sense caused by the failure of lighting. Mr. Chief Justice Burger's bicycle riding and drag-racing on Washington streets by teenagers are both causally linked to the accident the Chief Justice had. This is true regardless of the value society places on either riding bicycles or drag-racing on the public streets. It would remain true even if it could be shown that the particular injuries would have occurred even if the Chief Justice had been walking or even if there had been no drag-racers on the particular occasion. The point is that we believe that bicycle riding and drag-racing will increase the likelihood that such an accident will occur in the future even if on this specific occasion they played no role.[4]

Obviously there are an infinite number of acts or activities that are causally linked to every injury. Some may be viewed as socially desirable, some as neutral, and some as undesirable or even despicable. Getting up in the morning and going out increases the chances of many injuries—some of which are bound to occur. So does driving or practicing medicine. Driving or practicing medicine without a license increases the chances of injuries still more. It may be that

[4] I use the term "causal linkage" to describe situations where we *believe* that the occurrence of activity *A* will increase the chances of injury *B*. I do not mean by this to make any claims as to the existence of causal connections as they are sometimes described in philosophical writings.

we believe that these unlicensed activities increase only the chances of injuries due to negligent driver or doctor behavior, but it may also be that we believe they increase the chances of injuries in which no negligence can be found. In the first case, the activities are causally linked only to negligent harms; in the second, they are causally linked to a broader category of injuries, the incidence of which will increase whether or not negligent behavior occurs.[5] In sum, the concept of causal linkage between acts and activities and injuries is no more than an expression of empirically based belief that the act or activity in question will, if repeated in the future, increase the likelihood that the injury under consideration will also occur.

The second concept of cause is the familiar one of *but for* or *sine qua non* cause. A *but for* cause, as I shall use it, is any one of many acts or activities without which a particular injury would not have occurred. This usage is common enough and needs no special discussion. The only point worth making at this stage is that, while there is frequently an overlap between causal linkage and *but for* cause, the two concepts often diverge. The death of a person found with a broken neck at the foot of an unlighted staircase is causally linked to the absence of light even though it may be conclusively proved that the particular victim fell on the particular occasion because of a sudden dizzy spell that would have occurred with or without adequate lighting, that is, was not related in a *but for* sense to the absence of light. Conversely, *but for* the fact that the trolley driver had been speeding, the trolley car would not have been under a particular rotten tree when it fell and hit the car. Yet, unless speeding increases vibrations and vibrations increase the likelihood of trees falling, the admittedly *but for* cause would not be causally linked to the injury.

The final concept I shall consider is that of proximate cause. This concept, again familiar enough, will be used in a largely conclusory sense. It will embrace those presumably causally linked, usually (but by no means always) *but for* causes to which, in the absence of certain specified defenses, a particular legal system wishes to assign at least partial responsibility for an accident. Such a conclusory "definition" will obviously necessitate a discussion of what additional requirements, if any, must be met, under our current legal system and under each of the tort goals mentioned above, in order to conclude that an action is a proximate cause.

---

[5] *See, e.g.*, Brown v. Shyne, 242 N.Y. 176, 151 N.E. 197 (1926); *cf.* Ross v. Hartman, 139 F.2d 14 (D.C. Cir. 1943), *cert. denied*, 321 U.S. 790 (1944).

Why should tort law be concerned with causation in any of the three above senses of the word? To answer this question it is necessary to return to the goals or functions of torts I have discussed in earlier work and to ask how, if at all, causal linkage, *but for* cause, or the extra "something" that in practice is added by the requirement of "proximity" in cause may further those goals. As I have defined them elsewhere,[6] there are four such goals, two "compensation goals"—spreading and distributional equity—and two "deterrence goals"—specific or collective deterrence and general or market deterrence. In this paper I shall not attempt to reproduce those earlier definitions, but shall, instead, briefly explain what I mean by them as each one is examined in relation to the three concepts of causation outlined above.

## I. The Requirement of Causation in Torts: Cause as a Functional Concept

### A. Compensation Goals

Tort compensation goals may be described in rough terms as follows: since people not only value a lump sum of money differently, but place a different value on each subsequent dollar as well, the total impact of an injury may be diminished by an appropriate allocation of its burden. If, in general, a large lump sum burden is more onerous when borne by one person than it would be if divided among many, then one function of tort law may be to ensure that, consistent with other goals, injury burdens are spread. Moreover, if even heavy burdens are less onerous when borne by certain wealth categories rather than others, then another function of tort law, again consistent with other goals, may be to allocate injury burdens to those wealth categories able to bear them with relative ease. It is not necessary to consider here whether these are worthy goals; it is enough to note that they have played a significant role in tort law. For their significance, in itself, justifies asking the question: what role do the three concepts of causation have in furthering these goals?

1. *Spreading.* If spreading of injury losses were the only goal of tort law, there would be no point at all in requiring, as a prerequisite to liability, a causal link between an act or activity and the injury. The fact that some acts or activities increase the chances that an injury will occur in the future, while others do not, indicates

[6] G. Calabresi, *supra* note 2.

nothing about the ability of either injurer or victim to spread injury costs that have occurred. A prediction of the likelihood of future accidents resulting from particular behavior has no relation to ability to spread past injury costs. Hence, causal link between an action and an injury would be irrelevant to liability.

Of course, one could ask a different kind of causal link question, such as whether placing the loss on a particular actor would increase the chances that he would be able to spread similar losses occurring in the future. This kind of causal linkage would be fundamental to allocating losses so as to achieve spreading, but it has nothing to do with causation as it has been used in tort law.

All this is essentially tautological. Spreading is concerned with lessening the burdens of losses that have already occurred. Causal linkage, in its normal sense, refers to the likelihood that some behavior will result in loss in the future. Hence, causal linkage is irrelevant to the spreading function of tort law.

The same is true of the requirement of a *but for* relationship between an act or activity and the injury. The fact that some behavior was or was not a *but for* cause of an injury says nothing about the ability of those who engaged in that behavior to lessen the impact of the injury by spreading its burdens. Thus, any requirement that responsibility be accompanied by a *but for* link, just like any requirement that it be accompanied by a causal link, must find its justification in tort functions other than the spreading of injury losses.

If spreading were the only goal of tort law, a social insurance fund, raised through taxes assessed on a per capita basis, would be the optimal mechanism for achieving it. Such a compensation system, designed to spread all accident losses to the maximum degree possible, would not ask causal questions. Because ability to spread depends on factors completely separate from linkage to a particular accident (*but for* cause) or predictions of what future behavior is dangerous (causal link), this system would require neither causal link nor a *but for* relationship as a prerequisite to assessing the taxes used to compensate accident losses.

The requirement of proximity, however, may have some role to play in regard to spreading aims. If spreading were the only goal, there would be no need for a causal link or *but for* relationship; it would follow that there would also be no need for whatever additional limitations are imposed by the requirement of proximity. But if other tort goals, yet to be discussed, make the requirements of causal link or *but for* relationship desirable, then the additional requirement of proximity might be used to pick relatively good

"spreaders" from among those potentially liable actors who were, in a causal link and *but for* sense, "causes" of the harm.

To determine whether in practice the proximate cause requirement serves some spreading aims, it is necessary to look to the content of the term "proximate" as it is used by the courts and then to consider whether the requirement as thus applied helps to distinguish relatively good spreaders from poor ones. To the extent that such an exercise involves case analysis it will be dealt with below.[7] Some aspects of the requirement of proximity which clearly bear on the selection of good spreaders out of a mass of potentially responsible parties may, however, be briefly suggested here.

Some elements in proximate cause, like the requirement of foreseeability (in whatever form it takes and to whatever extent it is, in fact, deemed essential), are clearly germane to ability to spread losses. Though some commentators have argued that people, in practice, are apt to insure even against some harms they cannot foresee,[8] that argument is based on the vagaries of insurance company practices and presupposes that, apart from such practices, where a loss is not foreseeable, the chances that someone would insure against it are substantially less. Since spreading ability is thus crucially linked to insurance coverage, it is not hard to see how foresight, as an element in proximate cause, is itself linked to spreading ability.

Similarly, that aspect of proximate cause concerned with the "size and nature of the damages" is closely linked to the ability of the potentially liable party to "self-insure" (that is, either to set aside a contingency fund or to cover the losses out of ordinary income) in order to avoid being bankrupt by a large liability. It does not matter whether size and nature of damages are *explicitly* recognized as part of the concept of proximate cause. All that matters is whether, in practice, findings of proximate cause turn or appear to turn on the size of the damages and the ability of the defendant to bear them and still keep going. To the extent that they do, proximate cause furthers the goal of spreading.

Finally, the proximate cause requirement may also reflect spreading aims when it gives juries, instead of courts, the power to assess responsibility. If in determining who shall bear the loss juries are in practice more affected by ability to spread than are courts, then the requirement of proximity, by its very effect on who decides

---

[7] *See* text and notes at note 55 *infra*.

[8] *See* Morris, *Enterprise Liability and the Actuarial Process—The Insignificance of Foresight*, 70 YALE L.J. 554 (1961).

responsibility, may serve to select good spreaders from poor ones.

Again, one should recognize that the concept of proximate cause may at most serve as a limit, a device for selecting the best spreader out of a group of potentially liable actors who were originally identified on the basis of causation principles irrelevant to spreading. Proximate cause cannot be used to impose liability on a non-causally linked, non-*but for* cause that would be a highly effective spreader instead of on a less efficient spreader that is a causally linked, *but for* cause. By imposing some limits, the requirement of proximity may serve to introduce the goal of spreading into the causal determination; the requirement cannot, however, explain why causation should be an essential element of liability in tort. To repeat, a system of accident law designed with nothing but spreading in mind would be a social insurance system funded by general taxes; such a system would necessarily do away with any requirement of causation, including that of proximity. Thus, to find a justification for requirements of causation in torts, it is necessary to look beyond spreading goals.

2. *Wealth Distribution or "Deep Pocket."* For much the same reasons that spreading functions do not help explain the requirement of causation in torts, distributional goals also fail as explanatory principles. Again, causal linkage is not helpful in selecting from among potentially liable parties those who, if burdened, would better serve wealth distribution functions. The fact that certain activities increase the danger of future accidents says nothing about the relative wealth of the parties involved. The same is, of course, true of *but for* relationships. A *but for* cause of an injury is no more and no less likely to come from a wealthy category of injurer than a non-*but for* cause. Hence, the *but for* relationship is also irrelevant to wealth distribution goals.

Again, it is possible to redefine causal linkage and *but for* relationships so that they run, not from "action" to "injury," but from wealth, status, caste or even generalized merit (so long as "merit" is not defined in terms of accident avoidance[9]) to injury. Causal requirements could be shaped as follows: "in order to impose liability we must be able to say that, *but for* his wealth the injury would not have occurred" or "we require that his wealth increase the

[9] To the extent that accident avoidance is the basis of merit judgments and that these judgments in turn become the basis of distributional judgments, causal link and *but for* cause may, indeed, be relevant. But then the goal, in my terminology, is not a distributional one. Because its effect is to reward and punish for the purpose of diminishing accidents, it is, instead, a deterrence goal.

chances of his engaging in activities that in turn increase the probability of future occurrences of this injury." Such a "*but for*" or "causal link" requirement would, of course, serve distributional goals just as its analogue, requiring a *but for* relationship or a causal link to spreading ability, would further spreading goals. But, as noted above, tort law has never been concerned with these types of causal relationships. Rather, the *but for* relationships and causal links that have been at the core of tort cases are in themselves neutral insofar as spreading or wealth distribution are concerned.

The role of the proximity requirement is more ambiguous. Here too, since it serves only as a limitation, a device for selecting among potentially responsible parties, it cannot be used to pick out that party who, in distributive terms, is best suited to bear the burden. It may be used, however, to select a relatively good loss bearer in terms of wealth distribution goals from among a group of parties who are deemed potentially liable in terms of other tort goals.

It is more difficult to be sure of this role for the proximate cause requirement with respect to distributional goals than it was with respect to spreading, because some aspects of proximity (like foresight) relate directly to spreading capacity, while none go directly to distributional capacity. Still, it may well be that distributional considerations are implicit in how juries, and perhaps even courts, apply the test of proximity. In any event, there are enough indications that this may be the case[10] to suggest that, as with spreading, distributional goals may be furthered by the proximate cause requirement. Though again, as with spreading, if wealth distribution were the sole aim of tort law, such a requirement would be pointless.

## B. Deterrence Goals

Deterrence goals in tort law may be described as those which seek to minimize the sum of injury costs and safety costs. These goals are to be achieved not by mitigating the burden of costs that have already occurred, but by creating incentives so that people will avoid those future injuries worth avoiding and thus achieve an optimal trade-off between safety and injury in a world where safety is not a free good, and hence injury is not a total bad.[11] The object of

[10] *See, e.g.*, Kinderavich v. Palmer, 127 Conn. 85, 15 A.2d 83 (1940). *See also* Tedla v. Ellman, 280 N.Y. 124, 19 N.E.2d 987 (1939). In both these cases the courts' failure to find proximate cause as a matter of law is difficult to understand. One cannot conceive of the courts reaching the same results if the parties' relative financial positions had been reversed.

[11] What the optimal trade-off will be, of course, depends on what value is put on both injury and safety costs. Very different values may be placed on these factors, depending on

the following analysis is to examine the role that causal requirements play in achieving this trade-off and specifically to ask which requirements further which deterrence goals.

1. *Collective or Specific Deterrence.* One way of achieving the appropriate trade-off between safety and injury costs is through collective determinations of what acts or activities are too dangerous to be permitted. I have called this approach "collective" or "specific deterrence." In its pure form it involves a judgment by society that certain acts or activities are undesirable and are not to be tolerated regardless of individual desire to engage in them. In practice the high costs of enforcement may make total eradication of the disfavored behavior impossible. But whatever the cost limitations, the basic judgment remains the collective one to forbid the behavior.

If collective deterrence were the only goal of tort law, the requirement of causal linkage would play a crucial role. Society presumably decides to forbid or otherwise restrict certain behavior because it believes that such behavior is likely to increase the occurrence of injuries. The safety costs, that is, the burden imposed on those who must alter their behavior in order to diminish this risk, are deemed small enough to be worth imposing. The judgment involved in choosing to forbid or restrict particular behavior is precisely the kind of judgment that depends on the existence of a causal link, for there is no reason to prohibit or restrict behavior that we do not believe will increase the chances of injury in the future. Such behavior is, *by definition*, harmless to the best of our knowledge; since some people desire to engage in it, there is no reason why it should be barred.[12]

Obviously, not all behavior that is causally linked to harm is worth deterring collectively. Most behavior is risky and increases the chances of injury—that is why a trade-off between safety and injury costs is needed. Causal linkage to injury says nothing about the terms of the trade-off; it only defines those situations in which

---

the nature of the decisional mechanism chosen, whether it be a collective political process, a market process, or some combination of the two. A discussion of the choice of mechanism or even of the relationship between that choice and distributional goals is beyond the scope of this paper.

[12] Injury or harm cannot, of course, be limited to physical or economic damages. Behavior may be forbidden (whether wisely or not) because some people are "shocked" by it. Being shocked is being harmed and the "shocking" behavior is causally linked to the harm. I have elsewhere termed such harms "moralisms," *see* Calabresi & Melamed, *supra* note 2. The statement in the text takes no position on the appropriateness of restricting behavior that only offends. It simply states that behavior that is causally linked neither to offense nor to other injuries is, by definition, considered innocuous.

a trade-off may be needed. As such, the fact that an activity is causally linked[13] to an injury is a necessary but not a sufficient condition for the imposition of collective deterrence restrictions. Therefore, to define the prerequisites of liability under a collective deterrence standard it is necessary to look to other causation requirements.

In general, the requirement of a *but for* relationship between an action and an injury provides no guidance in selecting from among a universe of causally linked acts those which should be collectively deterred. If specific deterrence were the only goal of tort law, collectively proscribed behavior would be penalized regardless of whether in a specific instance it was a *but for* cause of harm. If drunken driving were forbidden because of its accident-causing potential, the drunken driver would be penalized whether he had been caught as a result of chance, of an accident, or of some other attention-getting behavior. Moreover, even if the driver were caught because he had been involved in an accident, a penalty would be imposed without reference to the specific cause of the accident; it would be irrelevant whether it was, in a *but for* sense, due to his drunken driving or whether it would have occurred even if he had been as sober as a law professor. The harm that seems likely to flow from drunken driving (the causal link) would support the penalty quite apart from any connection between the proscribed behavior and the specific accident that brought the behavior to our attention.

The requirement of a *but for* relationship might, however, further the goal of specific deterrence in those instances involving behavior collectively deemed worth proscribing only if the actor has a certain mental attitude, such as intent to injure or to risk injury. Distinguishing certain categories of behavior by the intent involved creates the difficult problem of determining when behavior, in itself permissible, bespeaks the intent that justifies the collective prohibition. Since intent to injure or to risk injury is not easily established, various surrogates (such as the "successful" completion of the injury) are used. It could be said that, while it is impossible to know that someone *actually* intended harm merely from the attempt, the inference of intent *is* sufficiently strong if he completed his attempt and injured someone. This line of reasoning, which is sometimes used to justify the imposition of different penalties for "attempted" and "completed" crimes, can also serve to explain some applications of a *but for* test in a world of pure collective de-

[13] Remember that causal link, as I have defined it, implies *belief*—not certainty.

terrence. For example, the collective judgment may be that only intentionally dangerous drunken driving must be forbidden; in that case "intentionally dangerous" could be defined as that drunken driving that results in an accident. Under such circumstances, if the injury would have occurred even without drunken driving, collective deterrence penalties might not be imposed on the drunken driver.

Even if the theories on which it depends are valid, such limited relevance would hardly explain the significance that *but for* causation has traditionally enjoyed in tort law. Moreover, even when applicable it would require a *but for* relationship only to the occurrence of an accident involving harm and not to the injuries produced by that accident. Proof of the intent necessary to proscribe the behavior would be provided simply by the occurrence of a harmful accident. The *extent* of the damages resulting from the accident would be irrelevant to the imposition of the collective deterrence sanction, unless the extent of the injury grew step-by-step as a function of the intent required to justify the sanctions. Apart from such rather bizarre situations in which the existence of the disfavored behavior could be determined only by looking at the precise degree of harm that occurred in a particular case, the *but for* test (as it is applied in tort law—that is, to damages, as well as to the accident itself) would seem to serve no function for collective deterrence purposes.

The general lack of relevance of a *but for* relationship to collective deterrence does not mean, however, that the requirement that the penalties assessed be related to foreseeable harm is also irrelevant. Although the notion, articulated in *The Wagon Mound*,[14] that the size of the penalty should depend on the size of the foreseeable future harm is occasionally confused with *but for* requirements, they are totally separate concepts. There need not be a *but for* relationship between penalized behavior and the injuries that in fact occur. Rather, the size of the penalty collectively deemed appropriate should be based on a notion of causal linkage refined to reflect the *extent* of those injuries that we believe may well be increased by the behavior to be sanctioned. Such a causal linkage requirement may or may not be appropriate to the kind of collective deterrence desired. It is enough now to note that *The Wagon Mound*, precisely because its reasoning sounds in collective deterrence, would, if carried to its logical end, do away with *but for* requirements. Wrongful behavior (defined in terms of its foreseeable propensity for harm)

[14] Overseas Tankship (U.K.) Ltd. v. Morts Dock & Eng'r Co. (The Wagon Mound), [1961] A.C. 388 (N.S.W.); *see* note 32 *infra* for a description of the case.

would be subject to penalties based on that propensity for harm, regardless of whether more or less harm than that foreseen actually came to pass—that is, regardless of *but for* relationships.

The requirement of proximity in causal relationships, unlike that of *but for* relationships, may help to select from the universe of causally linked actions those actions worth deterring collectively. It may do so in two ways.

First, the element of foreseeability in the requirement of proximate cause is directly relevant to collective deterrence. What, one may ask, is the use of trying to penalize or deter collectively those acts or activities whose propensity for harm cannot be known at the time the action takes place? After the accident a causal link may be recognized, because we now know that such acts or activities are dangerous, but why should that be a basis for penalizing actors who neither knew nor should have known of that risk before the accident? Such penalties cannot alter dangerous behavior. Thus, in terms of collective deterrence the argument for a foreseeability requirement excluding many causally linked actions from liability is very strong.[15]

The argument for foreseeability does not, however, require that the collective decision makers be able, *before the accident*, to define those actions that are foreseeably dangerous. Collective deterrence is as appropriate when the actors themselves should have foreseen the danger in time to control their behavior as when the collective decision makers prohibit it. Thus, open-ended categories of behavior, filled in after the action takes place, may be prohibited or penalized if it can be established at that time that the action was, in fact, dangerous and that the actor should have foreseen the danger before acting. Again, the actual occurrence, in a *but for* sense, of harmful injury would not be logically relevant although it might call the situation to our attention and thereby trigger the collective penalties.

Second, the requirement of proximity may be useful in deciding which causally linked activities should be selected for control out of the infinite number of possibilities, because proximate cause implies some look at the *relative* susceptibility of various actions to modifications diminishing their riskiness. It is true, as all first year torts students learn, that the requirement of proximate cause does not mean that only one among many causes must be determined to

[15] The analogous argument is much more tenuous if the goal is market deterrence. *See* p. 88 *infra*.

be proximate.[16] Yet the more subtle student also learns that whether a causally linked party is held responsible as a proximate cause depends, in part, on what other parties were involved and how risky their behavior was in comparison to his.

If the goal of collective deterrence is to select from activities that increase the risk of injury those worth controlling (in other words, if causal linkage suggests the need for a trade-off between injury and safety costs), then the requirement of proximity may serve a useful function by excluding from control some relatively less risky actions. One element in the trade-off between safety and injury costs in any particular activity is bound to be whether controlling or modifying another activity would be more efficient. The relational aspect in proximate cause seems to look to precisely that issue.[17]

This discussion of proximate cause as a device for selecting relatively good avoiders of accident costs for collective deterrence purposes may strike the reader as somewhat peculiar. After all, is that not, at least in part, what *fault* is about? Is fault not designed to select from an infinity of causally linked actions those which are *by definition* not worth doing, that is, wrongful or tortious? Is the balance between safety and injury that the existence of a causal link requires us to strike not the very same balance that sophisticated definers of fault from Terry and Learned Hand through Richard Posner say fault achieves?[18] The answer would seem, inescapably, to be "yes." And, indeed, if collective deterrence were the sole goal of tort law, it would be logical to reshape the definition of fault to

[16] This limitation on the extent to which the requirement of proximity selects from causally linked parties has sound bases linked to avoidance of administrative costs.

[17] This is not to say, however, that collective deterrence of only *one* activity is desirable. Considering which type of risky behavior should be forbidden does not mean that only one kind of risk-bearing behavior should be controlled for each injury. A class of injury may be best avoided by controlling several activities that are causally linked to the injury.

[18] Terry defined fault as conduct involving unreasonably great risk and isolated five factors (magnitude, principal object, collateral object, utility, and necessity of the risk) that could be used in determining whether a particular risk was so great as to be "unreasonable." Terry, *Negligence*, 29 HARV. L. REV. 40 (1915). Learned Hand in United States v. Carroll Towing Co., 159 F.2d 169 (2d Cir. 1947), proposed a formula for achieving the same result: the risk is unreasonable when the probability of harm ($P$) times the gravity of the injury that would result if the event did occur ($L$) is less than the burden of taking adequate precautions ($B$). He noted that, though the formula might suggest precision, its terms were in fact incapable of quantification. In Posner, *A Theory of Negligence*, 1 J. LEGAL STUDIES 29 (1972), Richard Posner interpreted the Hand formula in an explicitly economic way, equating $PL$ with the economic benefit anticipated from preventing accidents and $B$ with the cost of prevention. In utilizing this formula, Posner says, the law attempts to reach a "cost-justified" level of accidents.

do the whole job of selecting those causally linked acts and activities collectively deemed not worth doing. Because what is worthwhile doing necessarily depends on what alternatives exist for accomplishing the same result, such a refined fault principle would have to take into account what other ways were available for avoiding injury.[19] The size of the penalty to be imposed on faulty behavior would thus depend on the degree of control exercised over other risk-bearing activities. This concept, in a sense, was what the *Wagon Mound* court would seem, albeit very awkwardly, to have been groping toward.

If liability were explicitly determined in this manner, the notion of proximate cause would seem to become unnecessary in practice. Explicit collective consideration[20] of what activities are worth controlling relative both to their risk and to the risk incident to other activities which, if controlled, would also diminish injuries would be a satisfactory substitute for the requirement of proximity. Explicit determination of how large a penalty would best do the job of control, in view of the foreseeable risks of each activity, would similarly be more direct and effective than the roundabout way suggested by the language of *The Wagon Mound*. But the fact remains that the requirement of proximity has not yet been eliminated from tort law. After fault is found (after, that is, an action has been deemed not worth doing *because* of its riskiness and its relative avoidability), proximate cause must still be established. And penalties are not, *The Wagon Mound* notwithstanding, based on a collective judgment of what is appropriate in light of the foreseeable dangers of engaging in the controlled activity.

No wonder then that the requirement of proximate cause has seemed almost paradoxical in a fault system. Its very function as a "balance striker" between safety and injury costs casts doubt on the balance initially struck by the fault determination. What function does fault serve if the balance it strikes needs readjustment through a requirement of proximate causation? No ready answer can be found to this question if one views torts solely in terms of collective deterrence. For if that were the only goal of tort law, one could clearly do better by refining fault's balance than by playing the game according to the rules of proximate cause.

Thus, what was true of proximate cause and the compensation goals is also true of proximate cause and collective deterrence. Proximate cause, as applied, may well further these goals, but the goals

[19] This is, of course, what Posner's rather utopian definition of fault would do.

[20] If not by the judge, then by the jury as collective decision maker.

*by themselves* cannot serve to justify its use. To see in its full complexity the role that the concept of proximate cause plays in torts, as well as to understand, for the first time, the significance of the requirement of a *but for* relationship, we must turn to another goal of tort law—market deterrence.

2. *Market or General Deterrence.* Although it approaches the goal differently, market deterrence aims at the same object as collective deterrence. Both seek to strike a balance between safety and injury costs by avoiding only those injuries whose harm is sufficiently great to justify costly avoidance. But while collective deterrence would accomplish this end through a political, collective balancing of safety and injury costs, general or market deterrence would leave the judgment to an infinity of atomistic, individual market decisions.

Operating in a world in which perfect markets do not and cannot exist, a market deterrence approach places injury costs on those actors who can best decide whether avoidance is cheaper than bearing those costs. The object is to place the incentives for choosing between accident costs and their avoidance on those actors who, as a practical matter, we believe will choose most effectively. Thus, the chosen loss bearer must have better knowledge of the risks involved and of ways of avoiding them than alternate bearers; he must be in a better position to use that knowledge efficiently to choose the cheaper alternative; and finally he must be better placed to induce modifications in the behavior of others where such modification is the cheapest way to reduce the sum of accident and safety costs. The party who in practice best combines these not infrequently divergent attributes is the "cheapest cost avoider" of an accident who would be held responsible for the accident costs under the market deterrence standard.

If selection of such a cheapest cost avoider is the object of market deterrence, how helpful are the various concepts of causation I have distinguished for accomplishing it?

Generally a causal link between an activity and an injury would be required. It would clearly be unproductive to try to induce a modification in conduct for the purpose of reducing injury costs unless we believed the conduct to be causally linked to those injury costs. To put it another way, how can a person be the cheapest cost avoider of an injury if his actions do not increase the chances that the injury will occur? An incentive to modify behavior that appears to increase the chance of injuries is misplaced unless the party held

liable actually engages in such dangerous behavior.[21]

The role of *but for* causation in a system of market deterrence is less obvious than that of causal linkage. Nevertheless, market deterrence alone among the tort goals outlined above can explain the virtual universality of the *but for* test.

The function of the *but for* requirement in market deterrence is to assure that the injury costs allocated to the cheapest cost avoider include only those costs relevant to the choice between injury and safety. This function is performed by the creation of an actuarial basis from which actors can decide whether future safety costs are cheaper or more expensive than future injury costs. Which party should be the one to make this decision depends, in part, on the existence of a causal link, a prediction of future risk. But the future injury costs that should form the basis of the calculation of relative safety and injury costs for other future cheapest cost avoiders are defined by assigning to the loss bearer those past injury costs as to which it was a *but for* cause.

So viewed, the *but for* requirement, far from being the essential, almost categorical imperative it is sometimes described to be, is simply a useful way of toting up some of the costs the cheapest cost avoider should face in deciding whether avoidance is worthwhile. One could do away with the *but for* test and employ other methods to achieve the same end. For example, one could simply guess at the size of the injury costs that will be associated in the future with behavior causally linked to such injury costs.[22] But such an approach would be unnecessarily vague for a system of market deterrence. By using the *but for* requirement, we tell the chosen loss

[21] There is an exception to this requirement of causal linkage as a necessary, though not sufficient, condition for market deterrence allocation of injury costs. The exception, which concerns what I have termed the "best briber," is, however, more apparent than real. Logically, it is possible that an actor is so well-suited: (a) to undertake the cost-benefit analysis; (b) to determine what other activity could best be modified so as to pick the optimal combination of accident and safety-engendering behavior; and (c) to induce modifications in this second activity that we might wish to put liability and thus an incentive on this first "knowledgeable" actor even though it does not personally engage in behavior that increases the risks of injury. In such a bizarre case the best briber would be the optimal bearer of injury losses (from a market deterrence viewpoint) even though it would not be directly causally linked to the injury.

In practice, the problem of allocating losses to a briber who is not causally linked to injury would rarely arise. In a fault system, moreover, it would seem that it *cannot* arise since fault itself begins by requiring defendant behavior that could have been modified to avoid the injury. The issue has some marginal significance, however, to an understanding of causal language appropriate to liability systems that abolish the requirement of fault. For that reason it will be dealt with in some detail in the book of which this paper is only a part.

[22] In a way, that is what collective deterrence does when it guesses that certain behavior is to be proscribed because of its risk.

bearer that its burden will equal those costs that, *but for* its behavior, would not have been incurred; inevitably, therefore, we also tell the loss bearer that its future insurance premiums will be based on those injury costs that, in the same *but for* sense, have resulted from its past behavior. In this way we can approximate the optimal burden, that is, the burden that will create appropriate incentives to avoid injuries worth avoiding and not avoid those injuries that are too costly to eliminate.

Yet one may question whether the building up of actuarial insurance figures on a case-by-case basis through the use of the *but for* test is the most efficient way of determining the burden to be placed on the cheapest cost avoider. Random samples of injury costs associated with certain types of behavior might form an equally precise and far less expensive way of setting up the incentive for a correct cost-benefit analysis. The proper choice of method in this instance depends on whether the administrative costs of case-by-case determination are viewed as fixed, that is, *already* incurred in order to decide *whom* to burden or whether they are viewed as marginal, that is, worth bearing only insofar as they give us a more accurate calculation of what costs should be placed on an already selected cheapest cost avoider. If we continue to use case-by-case determination to decide *who* should bear the loss, then a *but for* requirement to define the loss to be borne is not administratively expensive. If, instead, the category of actor picked to bear the loss is determined on a more general basis (say, by legislation), then a sampling of injury costs that could have been avoided by appropriate behavior modification might well be a more efficient way of calculating the size of the burden (it would be called a tax, rather than damages) to be imposed. Modification of the loss bearer's behavior would reduce the injury costs in future samples and this modification would, in turn, reduce the future tax burden. Since to date case-by-case determination has been taken for granted as an essential mechanism for deciding who should bear losses, however, it is little wonder that the *but for* test should also have seemed an essential part of the system.

To say this, however, is not to say that the *but for* test should be viewed as an absolute requirement in case-by-case determinations. Thus, where it is difficult to prove a *but for* relationship[23] or where either of two independent defendants was a suffi-

[23] *See, e.g.*, Michie v. Great Lakes Steel, 495 F.2d 213 (6th Cir.), *cert. denied*, 419 U.S. 997 (1974), holding that a group of polluters could be liable even though the harm attributable to each could not be shown; Summers v. Tice, 33 Cal. 2d 80, 199 P.2d 1 (1948), where the

cient cause of the harm and hence neither was a necessary (*but for*) cause,[24] it is, at the very least, doubtful whether blind adherence to the requirement that the victim prove a *but for* relationship serves the purposes of market deterrence. The fact that the victim's behavior is also a *but for* cause in such cases is likely to be viewed as irrelevant if the victim is not a desirable bearer of the burden in terms of market deterrence. Functionally, the issue is this: should the loss be allocated to the defendant in defiance of the strict *but for* test or is it better to let the loss fall on the wrong party, the victim, who by definition is not the cheapest cost avoider? It is little wonder that in cases of this sort courts have striven mightily, in the face of the seemingly inexorable requirement of *but for* relationship, to ignore the prerequisite and place the loss correctly, if inexactly, on the defendants.[25] The limited significance of the *but for* test in furthering the only goal of tort law that it serves would demand no less.

General or market deterrence, then, relies on causal linkage as one crucial element in identifying the optimal loss bearer. It relies on *but for* cause as a sound way of determining, on a case-by-case basis, what burden-incentive should be placed on the loss bearer. Yet neither concept, together or separately, suffices to identify the proper loss bearer. The requirement of proximate cause is necessary to select from actors who may be cheapest cost avoiders because they are sufficiently causally linked, those who in fact are.

Various elements in the requirement of proximate cause are relevant to selecting the cheapest cost avoider. Foreseeability is obviously germane, for clearly the ability to foresee risks is important in comparing accident avoidance costs with safety costs. It may seem strange that under a proximate cause test costs are allocated on the basis of *past* foreseeability, since under a market deterrence rationale such allocations are designed only to affect *future* choices between safety and accident costs. Inquiry into future risk—that is,

---

plaintiff had clearly been shot by one of two defendants, but it was impossible to tell which one; the court upheld the jury verdict for the plaintiff by explicitly shifting the burden of proof on the question of causation to the defendants; Paine v. Gamble Stores, Inc., 202 Minn. 462, 279 N.W. 257 (1938), where plaintiff's deceased was found at the bottom of a pit and the top rail around the pit was missing; the court affirmed a verdict for the plaintiff, apparently holding that once the plaintiff had shown that a defective railing *could* have been the *but for* cause of the harm, the burden of proving that it was *not* the cause shifted to defendants.

[24] *See, e.g.*, Corey v. Havener, 182 Mass. 250, 65 N.E. 69 (1902), where a pair of independent motorcyclists were held liable for frightening a horse, although it appeared likely that each cyclist alone would have made enough noise to bring about the harm.

[25] This result will generally be reached unless it seems desirable to put an incentive on the victim to identify which defendant was actually the *but for* cause of the harm.

degree of causal linkage—rather than into past foreseeability would seem appropriate. Yet it is probable that parties who have had relatively good information about possible risks in the past (that is, were relatively good foreseers) would also have such information about the future. Moreover, allocation of those costs, which were foreseeable in the past, will create incentives for both the loss bearers and others to foresee those injuries that may be worth avoiding in the future. Thus it follows that, as a practical matter, past foreseeability is a useful guide to finding the cheapest cost avoider.

Past foreseeability is not, however, a prerequisite to liability if the object is market deterrence. A cheapest cost avoider of future costs exists even though no one could have foreseen what was to transpire. Even though past injuries were not foreseeable, placing liability for them on one party rather than another may beneficially affect future choices from a market deterrence standpoint. If the effect of such allocations is simply to categorize certain activities as especially prone to unforeseeable and uninsurable risks, a favorable market deterrence effect may be achieved.[26]

Still, foreseeability as it is employed in proximate cause remains a useful factor for defining those activities that market deterrence would burden with accident costs. Moreover, the foreseeability of different *aspects* of an injury may produce a variety of gradations in ability to choose between injury costs and safety. In fact, this is how the courts use foreseeability in testing for proximate cause; as we shall see, the classic rules distinguishing requirements of foreseeability of type of risk and foreseeability of category of plaintiff from requirements of foreseeability of extent of damages make a fair amount of sense in a rough-and-ready way, if a significant object of the exercise is allocation of losses to the cheapest cost avoider.

Other aspects of proximate cause are equally relevant to the determination of which activities are best burdened as cheapest cost avoiders. Whether one category of loss bearer rather than another is the best arbiter of future safety and accident costs, or whether both are equally competent, depends on many characteristics of the activities viewed in relation to each other. Distance in time and space or the existence of intervening wrongdoers (or simply of intervening doers) may or may not make a particular actor unsuitable as a market deterrence loss bearer—even though it was at fault and even though the actor that is the better loss bearer (whether it be the victim or the injurer) was not.

---

[26] *See* p. 93 *infra*.

The very fact that these factors, crucial to proximate cause analysis, are not absolute (existence of an intervening cause may, but need not, make a previous cause too distant) renders them especially useful in furthering market deterrence. Under market deterrence principles, distance in time and space or the existence of intervening causes may make a party engaging in a causally linked activity so clearly less suited to choose between safety and accidents as to make burdening it undesirable. On the other hand, these factors may be irrelevant if the "nearer" actors are less knowledgeable or are, for some other reason, able to shift (externalize) the loss and hence lack the incentive to choose. Similarly, the fact that one actor is at fault (that is, we can collectively say that it ought to have avoided the accident) may or may not mean that it is a better future cost avoider than another actor who, because more proximate, has greater knowledge of what its choices entail. And this uncertainty persists whether or not we can collectively say that this second actor *ought* to have chosen safety (that is, was at fault). Proximate cause limits the degree to which even a fault system allocates losses on the basis of fault by considering as relevant, but not conclusive, factors that help define the cheapest cost avoiders. In this way it serves the goal of market deterrence, which would also consider those factors as relevant, but not decisive, and which would not be overly impressed by the collective judgment implicit in a finding of fault.

Some of the factors relevant to selecting the cheapest cost avoiders could be introduced by doctrines other than proximate cause. Causal linkage, for example, could be useful for this purpose, especially if one considered degree of causal linkage (that is, predictability as to how much the recurrence of one activity, relative to others, increased the risk of injury). But degree of causal linkage by itself would not raise all those issues that "proximity," as it is broadly understood, raises relevant to defining the cheapest cost avoiders. It does not, for example, take into account the relative knowledge of the danger possessed by the different parties. Another doctrine, assumption of risk, can readily take that into account, but it, in turn, does not focus on the capacity of the parties to externalize the costs and thereby neutralize any incentive effect. Proximate cause, by its very vagueness, permits these and other factors to be considered. It is likely, therefore, to remain significant even if the role of other doctrines in furthering market deterrence were expanded.

Proximate cause, though siguificant in furthering market deterrence, cannot be fully explained in terms of market deterrence, nor can it do the whole job required by market deterrence. While

the doctrine may serve to limit the liability of faulty injurers and leave the losses on non-faulty, perhaps more proximate victims, it cannot serve to do the reverse and burden faultless, but "very proximate" injurers. And, if the sole object of tort law were market deterrence, more direct ways than proximate cause could be discovered for applying the factors relevant to selecting the best loss bearer. But tort law does not serve a single goal. There are elements in the requirement of proximate cause that are useful in furthering the compensation and even the collective deterrence goals of tort law. As such, even though it would not serve any single goal ideally, the requirement may be fundamental to our actual system of tort law based, as it is, on many goals.

The same is, of course, true of the other requirements of causation. The deterrence goals generally require causal linkage; and they may be so fundamental in tort law as to make that requirement an almost absolute prerequisite to liability. Given case-by-case determination of responsibility, market deterrence may add *but for* cause as a useful, indeed almost fundamental, prerequisite. At that point the requirement of proximity in causation, although potentially useful in the achievement of both collective and market deterrence, may instead be employed to further spreading. For example, imagine an accident unforeseeable to either victim or injurer, in which, after the accident, we find that the behavior of both injurer and victim was a *but for* cause and was causally linked to the injury.[27] In such a case, whether the losses were left on the victim or moved to the injurer might depend on their relative ability to choose between injury and safety in the future, or it might depend on which party could best spread the losses already incurred. As a practical matter, the test of proximate cause could be used to do either. Indeed, it could serve one goal in one such case and the other in a different one. Nearly as good cost avoiders might have very different spreading potential in one case, while equally good spreaders might be very differently situated to choose between injury and safety in another.

Causation, viewed as a set of functional concepts, must respond to the sum of the goals of tort law. Causal linkage and *but for* cause are functionally related to few, but fundamental tort aims. Proximate cause is, however, related to all such goals. It is little wonder then that it has always seemed the most complex, paradoxical, but also "fiexible" and policy-based of the causal requirements. Indeed,

[27] Both parties' behavior may have been faulty or botb may have been blameless or one faulty and the other blameless.

its very flexibility and explicitly functional policy orientation sometimes serve to hide the equally functional, rather than absolute, roles played by the other concepts.

## II. Application of the Model to a Few Traditional Cases

The foregoing analysis of causation in terms of functional tort goals may be applied in determining liability under both fault and non-fault systems. Given the present space limitations, however, I shall apply it here only to a few classic cases decided within the fault system under the rubric of "proximate cause" and the subsidiary concept of foreseeability of harm.[28]

### A. Market Deterrence and Foreseeability of Unusual Damage

As noted above,[29] use of the concept of *but for* cause in addition to the concept of causal link can be understood only insofar as one goal of tort law is market deterrence. Both causal elements have traditionally been accepted as virtually absolute prerequisites to liability under the fault system, suggesting that in practice the law has considered market deterrence to be a crucial goal of tort law. Taking as given that market deterrence is a fundamental starting point of causal analysis and assuming that the prerequisites of market deterrence (causal linkage and, to a lesser extent, *but for* relationship) have been satisfied, has the law utilized the requirement of proximate cause in a manner consistent with the promotion of market deterrence goals? An exploration of three classic situations dealing with one element of proximate cause—foreseeability of unusual damage—should serve as a convenient vehicle for examining this question.

The first case, exemplified by *Palsgraf v. Long Island R.R.*,[30] involves limited or absent foreseeability of category of plaintiff. In *Palsgraf* railroad employees helping a passenger onto a train negligently caused him to drop a package he was carrying. Unknown to them, the package contained fireworks. When it fell, the fireworks exploded, causing some scales on the other side of the platform to fall and strike Mrs. Palsgraf, injuring her. While harm to the passenger boarding the train was a foreseeable risk of the negligent

[28] In the forthcoming book I will apply the analysis to many other, more difficult "fault" cases and to various systems of non-fault liability.

[29] *See* p. 85 *supra*.

[30] 248 N.Y. 339, 162 N.E. 99 (1928).

behavior, harm to Mrs. Palsgraf, who was standing well out of the way, was held not to be. At most it was only barely foreseeable that endangering the package might harm a distant bystander.[31]

The second case, epitomized by the "thin skull" or "eggshell plaintiff" hypothetical, involves limited or absent foreseeability of extent of damages. The defendant does something wrong. It is deemed faulty because it is easily avoidable and likely to do the plaintiff a small, but not insignificant injury. Plaintiff has a thin skull and dies. Plaintiffs with thin skulls *are* foreseeable (whatever courts might say), but their presence (like the chance of harm to bystanders in *Palsgraf*) is sufficiently rare that the foreseeable harm would not, by itself, justify calling the defendant's conduct negligent.

The third case involves limited or absent foreseeability of what is rather loosely termed "type" of damage. The defendant gives his three-year-old daughter a loaded gun; the daughter drops the gun on the toe of an elderly relative. Even though the gnn does not go off, the relative is seriously injured. Loaded guns are very difficult objects for three-year-olds to hold; this fact, which establishes causal linkage, is either unforeseeable or not sufficiently likely to make handing loaded guns to infants faulty behavior. What makes the conduct faulty is, of course, the danger that the child will fire the gun; but that did not occur.[32]

What effect, in all these cases, should limited foreseeability or absence of foreseeability have on the proximate cause decision when that decision is viewed from a market deterrence standpoint? The issue posed by this question would seem to be no more than: which category, plaintiff or defendant, had the better ability to choose the safer alternative as far as this unexpected or barely expected damage was concerned?

---

[31] By stretching the term "foreseeable," it would be possible to construct a scenario in which Mrs. Palsgraf would have been a foreseeable plaintiff. Thus, the first passenger might have attempted to retrieve his package, despite the fact that the train was moving. In doing so he might bave been injured. Mrs. Palsgraf and various other bystanders might have rushed up to help (after all, danger invites rescue). And in the rush, Mrs. Palsgraf might have been hurt. Nevertheless, this danger of injury to Mrs. Palsgraf would not, *by itself*, have justified calling defendant's conduct wrongful.

[32] Overseas Tankship (U.K.) Ltd. v. Morts Dock & Eng'r Co. (The Wagon Mound), [1961] A.C. 388 (N.S.W.), is another example of a case in which the type of damage was unforeseeable. In that case the defendant's employees had negligently discharged furnace oil into Sydney's harbor, thus risking some foreseeable harm by the oil as grease or muck. However, plaintiff's workmen, who were using acetylene torches, accidently ignited some rags which in turn ignited the furnace oil, causing extensive damage in the harbor. Because the defendant could not have foreseen the risk of fire the Privy Council refused to hold him liable.

The fact that one party cannot foresee a certain kind of damage does not automatically exclude it as a cheapest cost avoider, although the chances that it is are reduced. Unforeseeable risks may, as a statistical matter, cluster around certain activities. If those engaging in such activities are held liable for those unforeseeable damages, the activities will acquire a reputation for being financially risky. As such, they may be undertaken to a lesser extent or undertaken only by those who find them especially valuable, by people who, more than most of us, enjoy taking risks, or finally by entrepreneurs, that is, those who are willing to gamble on their knack for avoiding even unforeseeable risks more successfully than most people. Thus, even in the absence of foreseeability or in the presence of only very limited foreseeability, such modifications would serve to reduce the sum of accident and safety costs, that is, to further market deterrence. Indeed, they may do so better than allocation of the risks to the injured plaintiff would, even if the plaintiff could foresee the very "excess" risks that we have assumed the defendant category could perceive barely, if at all. The plaintiff category might, if burdened, shift the loss to another category, such as a social insurance fund, which would be totally incapable of choosing between safety and accident costs. Or the plaintiff category might be unable to modify its behavior even if it foresaw the risks of the excess injury. Still, the less the risks are foreseeable to the defendant and the more they are foreseeable to the plaintiff category, the more plausible it is that the plaintiff category is the cheapest avoider of the costs involved.

Given all this, one can begin to understand how the distinctions between barely foreseeable or unforeseeable category of plaintiff (*Palsgraf*), extent of damages (the thin skull case), and type of damage (the loaded gun case) may, in a very crude sense, correspond to distinctions made in a search for the cheapest cost avoider.

1. *Foreseeability of Category of Plaintiff.* The fact that we are prepared to say that the defendant could have avoided a given loss more easily than one category of plaintiff says very little about the relative ability of the same defendant to avoid the same loss when compared to a totally different and perhaps unforeseeable category of plaintiff. The first category of plaintiff might not be able to foresee the loss at all or might not be able to act on whatever he could foresee. On the other hand, a different plaintiff category might be in a much better position than the defendant to foresee the loss and balance its risk against safety costs. If the second category of plaintiff is unforeseeable, the defendant may be unable to gauge whether this second plaintiff was a better arbiter than the defendant as to

safety and injury costs and therefore was the one who should have undertaken whatever safety measures, if any, were appropriate to avoid injury to himself. While this ignorance does not absolutely exclude defendant liability, it does make it considerably less likely, for one aspect of being the cheapest cost avoider is knowing that you are likely to bear the burden of making the cost-benefit analysis.

At the very least, the existence of an unforeseeable or only slightly foreseeable plaintiff requires the court or jury to make an independent examination of relative cost avoidance potential between the defendant and this plaintiff category. The fact that the defendant was the better cost avoider vis-à-vis another foreseeable plaintiff category would tell us little, if anything, about how this examination should come out. This is true even when different categories of *foreseeable* plaintiffs are involved, but we would expect differences in result to become more pronounced as the categories of plaintiffs become increasingly unexpected. For example, the fact that a defendant landowner is the cheapest avoider of certain losses with respect to most categories of invitees does not mean that he is also the cheapest cost avoider with respect to many categories of trespassers or even certain licensees.[33] Similarly, the fact that a construction company is the cheapest avoider of injuries to children caused by loose pipes left on a playground may, but need not, mean that it is also the cheapest avoider of injuries to adult fools who are also injured there.[34] The issue in the latter instance is whether the category "fool" is a cheaper avoider than the category "construction company." In the given case I doubt that it would be, and I expect that a court would agree and impose liability. However, the fact that the construction company was a better cost avoider than the category "children" would not be decisive.

The foregoing analysis leads to the conclusion that Cardozo's emphasis on category of plaintiff in *Palsgraf*, though too strong if taken literally (as Cardozo in fact never did), is understandable largely, if not totally, in terms of market deterrence. As *The Wagon Mound* in effect demonstrated, it is not comprehensible in terms of collective deterrence. What the defendant did was wrongful and to be avoided because of foreseeable harm to one kind of plaintiff, so how can it be any less wrongful because harm occurred to another category of plaintiff? Probably, as *The Wagon Mound* suggests, collective deterrence would require a limitation of the penalty to the

[33] I do not mean to suggest, however, that the virtually discredited "status on the land" categories are sufficiently refined to provide an adequate test.

[34] *See, e.g.*, Terranella v. Union Bldg. & Constr. Co., 3 N.J. 443, 70 A.2d 753 (1950).

foreseeable *degree* of harm, but surely it would be irrelevant to collective deterrence whom that foreseeable extent of harm struck.[35] Paradoxically, given that he never openly questioned the premises of the fault system and therefore, in a sense, the goal of collective deterrence, Cardozo's approach can only be explained in terms of market deterrence or compensation goals. Significantly, however, these goals, though only implicit, were often readily identifiable in his opinions.[36]

2. *Foreseeability of Extent of Damage.* The very crude way in which limited or absent foreseeability of category of plaintiff is relevant to finding a cheapest cost avoider has little parallel when the limited or absent foreseeability relates simply to the extent of damages that are of the same "type" as those readily foreseeable and that injure the same category of plaintiff. By and large, if a defendant category is the cheapest avoider of one type of damage vis-à-vis a given category of plaintiff, then that same defendant will be the cheapest avoider of any related, but unforeseeable or barely foreseeable extra damages to the same category of plaintiff. While it is possible, it is not very likely that a plaintiff category that was not as well-suited as the defendant to avoid the readily foreseeable harm would be better suited to avoid the excess damages. Hence, the general common law rule that, once proximate cause between defendant's fault and the injury exists, liability extends to more serious, but unexpected damages as well, seems to be a good starting point from a market deterrence point of view:[37] It is, however, only a starting point, a rough approximation. Like the rule that liability exists only to foreseeable plaintiffs, it makes sense in terms of mar-

[35] The same result would follow from a strict reading of the Learned Hand test. The "avoidance" side of the balance holds nothing: the defendant is already bound to alter his behavior to prevent injury to the fully foreseeable plaintiff, hence there is no further avoidance cost implied in protecting against other possible plaintiffs. The danger side of the balance holds whatever peppercorn of risk the miniscule foreseeability of these other categories of plaintiffs entails. Infinitesimal though the foreseeability is, however, it is enough to tilt the scale, since there is nothing on the other side. Thus, under the Hand test, the defendant must be held negligent to the barely foreseeable plaintiffs, if he is negligent to some highly foreseeable ones.

[36] For a Cardozo opinion in which the cost-benefit analysis involved in accident cases was explicit, see Adams v. Bullock, 227 N.Y. 208, 125 N.E. 93 (1919). The effect of compensation goals may help explain the difference between his opinions in H.R. Moch Co. v. Rensselaer Water Co., 247 N.Y. 160, 159 N.E. 896 (1928), and MacPherson v. Buick Motor Co., 217 N.Y. 382, 111 N.E. 1050 (1916).

[37] Collective deterrence, and hence the *Wagon Mound* rationale, would, of course, require a different tack. Liability would be limited to the foreseeable extent of harm. Yet, if applied fully, this rationale would require the same level of liability even when that degree of harm failed to materialize.

ket deterrence only if, in practice, exceptions are made to it. The Cardozo "foreseeability of plaintiff" rule works reasonably well because the content of the term "foreseeability" is sufficiently flexible and manipulable to permit appropriate deviations. Similarly, the rule imposing liability for unlimited, unforeseeable damages works reasonably well, from a market deterrence standpoint, because its effect is controlled by other doctrines which, in appropriate cases, may be used to reduce its scope.

The doctrines used to limit the scope of the rule permitting unforeseeable damages *ad infinitum* are not, by and large, proximate cause doctrines.[38] They tend instead to be introduced through concepts of assumption of risk (in its primary sense[39]) and, on occasion, through contributory negligence. A great violinist who goes to work in a steel mill and whose hands are mangled as a result of the mill owner's fault is not likely to recover for his "extra, unforeseeable" damages. Many courts would say that he had assumed the risk of such harm, because he had greater knowledge of the unusual possible damages and could have found alternate employment less dangerous to his hands. Yet the violinist would probably be able to recover for the same excess damage to his hands if they were injured in a car accident. He might have greater knowledge of the particular risk involved than the defendant could have, but that knowledge would not provide him with any meaningful alternatives.[40] The defendant in this situation, unlike the defendant in the steel mill case, would, if he were the cheapest avoider of the original injury, also be the cheapest avoider of the unlikely extra damages.

The same types of distinctions that limit and refine the rule permitting unforeseeable damages so that it serves market deter-

[38] By defining the excess damages as being of a different "type" from those that have been foreseen, courts can, of course, employ proximate cause doctrines to limit recoveries of unforeseen or barely foreseen excess damages. Andrews, J., dissenting in *Palsgraf*, would bave employed proximate cause directly to limit the extent of damages, but foreseeability in his view would have been only one factor in determining the appropriate limits.

[39] In its primary sense, assumption of risk does not require that the plaintiff's conduct be in any way negligent or wrongful. It does require that the risk in question is one of which people like the plaintiff are ordinarily aware, that the plaintiff bas meaningful alternatives to bearing the risk, and that the defendant has a right (a weasel word) to impose the choice between taking and avoiding the risk on the plaintiff. *See, e.g.*, Brown v. San Francisco Ball Club, Inc., 99 Cal. App. 2d 484, 222 P.2d 19 (1950); *cf.* Meistrich v. Casino Arena Attractions, Inc., 31 N.J. 44, 155 A.2d 90 (1959), which rejects assumption of risk in its secondary sense (as a sort of contributory negligence) and questions whether its primary function is best accomplished by a separate doctrine, like assumption of risk, or through definitions of concepts like negligence and duty designed to achieve the same kinds of limits on liability.

[40] In our society staying at home is not a meaningful alternative, and protective gloves do not exist.

rence goals can be seen in other classic hypotheticals as well. If the package belonging to the passenger in *Palsgraf* had contained a Ming vase, the defendant probably would not have been held liable for the extraordinary damages caused by the breakage, despite the basic common law rule imposing liability and despite the fact that the risk that made the defendant's behavior faulty in the first place was precisely the risk of damage to the package.[41] Some courts would have found contributory negligence in this situation, others assumption of risk; whatever the terminology, the result appears foreordained and in accord with market deterrence goals. Conversely, the person with a thin skull generally recovers for all his injuries even though his knowledge of the condition must be vastly superior to the defendant's. Knowledge without alternatives (adequate helmets do not exist; if they did, the cases might well be different) does not make a category of actor a cheap cost avoider. And placing liability on the plaintiff as a market deterrence incentive to *inform* potential defendants of the particular risk seems relatively useless in such cases since the defendant, by and large, will be a driver who could not readily receive the information. In contrast there are cases in which information *can* be given, as where signs are erected informing drivers that a nursing home is nearby or, to use the reverse case, where *adequate* warning of potential harms from the use of a product is given and the burden of making the cost-benefit analysis is thereby shifted to the plaintiff, who can no longer successfully claim that the harm resulted from a product defect.[42]

Such non-proximate cause limitations on the common law rule governing excess damages are, of course, fairly crude. To produce more refined rules, it would be necessary to refer to market deterrence directly as the *only* goal of tort law, an approach that would clearly be incorrect. In all probability such a direct reliance on market deterrence would limit "extra" or "unusual" damages more than is done by existing tort doctrines. For example, it has been suggested that at least highly individualized pain and suffering might be left on the victims of some categories of injuries, even if other damages, including some basic level of pain and suffering, were placed on injurers.[43] The tone of such writings suggests market

---

[41] *Cf.* the last paragraph of Cardozo's opinion in *Palsgraf*.

[42] This definition of the functional role of warning and defect is reflected in Sills v. Massey-Ferguson, Inc., 296 F. Supp. 776 (N.D. Ind. 1969). *See generally* Henningsen v. Bloomfield Motors, Inc., 32 N.J. 358, 161 A.2d 69 (1960).

[43] *See, e.g.*, HALDI ASSOCIATES, INC., A STUDY OF HAWAII'S MOTOR VEHICLE INSURANCE

deterrence reasons for this split allocation. Some legislation, generally of a non-fault, market deterrence sort, has followed this model.[44] All that, however, begins to push us outside the scope of the present discussion toward a consideration of the role of causal concepts in a world of non-fault liability where one might plausibly argue that the specific goals of tort law should be clearly identified and served.

3. *Foreseeability of Type of Damage.* Ability to foresee category of plaintiff, as the common law defines it, is highly relevant to selecting cheapest cost avoiders. Ability to foresee extent of damages is much less relevant, especially given the existence of concepts like assumption of risk. Ability to foresee type of damages, once more as defined by the common law, lies somewhere in between.

The example used above for unforeseeable or only slightly foreseeable *type* of injuries was the case in which an adult handed a loaded gun to an infant. The child drops it on the toe of an elderly person who is seriously injured even though the gun does not go off. It is known, either before or after the accident, that loaded guns are particularly hard for infants to hold. Causal linkage is, therefore, established because the evidence indicates that handing loaded guns to children substantially increases the risk that they will be dropped and thus the risk that elderly feet will be injured as a result.

Our initial, instinctive reaction to this rather outlandish hypothetical is to want to hold the negligent defendant liable. That reaction, I would suggest, is based on collective deterrence grounds. The foreseeable harm from handing an infant a loaded gun is enormous; the harm that occurred is certainly no greater than the expected harm; the collective deterrence penalty appropriate to the defendant would be at least as great as the damages suffered; then why not hold the defendant liable? Such reasoning, of course, would justify the same serious penalty even if *no* harm had occurred. In a system of mixed goals, penalizing dangerous behavior in the absence of actual harm might be inappropriate, and yet collective deterrence goals might still compel the imposition of liability in the hypothetical case if, at the very least, other goals were not clearly furthered by non-liability.

---

PROGRAM (Special Report No. 72-1, to the Legislature of the State of Hawaii) 90, 110 (1972); *cf.* G. CALABRESI, *supra* note 2, at 223-24.

[44] Leaving pain and suffering damages on the victims was one of the major proposals in most compensation schemes for victims of automobile accidents. *See, e.g.*, R. KEETON & J. O'CONNELL, BASIC PROTECTION FOR THE TRAFFIC VICTIM: A BLUEPRINT FOR REFORMING AUTOMOBILE INSURANCE (1965). In addition to automobile compensation statutes, workmen's compensation statutes have also declined to impose highly individualized pain and suffering damages on the injurer.

This collective deterrence effect may be demonstrated by contrasting a variation on *Palsgraf* with the hypothetical case. In *Palsgraf* the defendant's negligence consisted of unnecessarily endangering the package. Suppose that the owner of the package had had a "thin shoulder" and that, while shoving him onto the train in a non-negligent manner except with respect to the package, the railroad employees jostled his arm so that it was severely injured and had to be amputated. Here, unlike the gun hypothetical, there is no collective deterrence equivalence between the wrong committed and the penalty to be imposed; hence, our instinct is to say "no liability." Cardozo himself suggested that result in *Palsgraf*. Yet that same result does not commend itself in the ordinary thin skull case where equivalence between wrong and penalty is also missing.[45] In the thin skull case damage is of the same type as that foreseen and thus market deterrence (absent assumption of risk) is furthered by placing the loss on the defendant despite the lack of equivalence between his wrong and the burden imposed on him. The above variation on *Palsgraf*, in which the type of damage is different, is less clearly a case where the negligent defendant, though the cheapest avoider of the easily foreseeable risk, is generally the cheapest avoider of the less foreseeable one as well. In such a situation the courts, if guided by market deterrence principles, would hesitate to start with a general rule of liability that was limited only by other doctrines such as assumption of risk. But they would be equally hesitant to start with a rule of non-liability (as in unforeseeable plaintiff situations) to be modified only through manipulation of the content of "foreseeability." It is little wonder, then, that the pressure of other goals, such as collective deterrence, is more obviously felt in this case.

The circumstances surrounding the occurrence of an unexpected type of damage may suggest that the defendant who is the best bearer of the basic loss for market deterrence purposes is not the best bearer of the different type of loss. But they may suggest the opposite. Or they may simply leave us perplexed on the point and hence more prone to follow the dictates of other goals (whether collective deterrence as in the "loaded gun" case or compensation), which may seem to be clearly furthered by one result or another. It is not surprising, therefore, that such cases appear at first glance to be the most mixed up of all the "unexpected damage" cases. Courts

[45] The exception is, of course, *The Wagon Mound*, which is not followed consistently *precisely* because its logic is entirely collective deterrence, and torts is not based on that goal alone; *cf.* Smith v. Leech Brain & Co., [1962] 2 Q.B. 405.

and commentators talk one way, or occasionally both ways, but however they talk, "type" is defined so loosely that cases come out every which way.[46]

This very confusion would seem to suggest that it would be better to concentrate directly on the effect that foreseeability of either extent, type, or plaintiff has on relative cost avoidance potential. Would it not be better to lump all such cases under proximate cause and then consider, given the particular facts, which category of actor was the cheapest cost avoider? If, as I suggest, the different foreseeablities are at best only suggestive of cheapness of cost avoidance and if market deterrence were the sole goal of tort law, it would *seem* better to treat all of them as raising the identical problem—how does relative forseeability affect market deterrence in this instance?—and then find the solution according to the facts of the particular case. All the difficulties of categorization that have long plagued the courts—is this a case of different type or extent of damages?—would then be avoided.[47] Nevertheless, the suggested solution does not necessarily follow. Focusing clearly on market deterrence may, in practice, fail to further even market deterrence goals.[48] And it would likely alter the treatment currently accorded other goals of tort law, goals that are no less important than market deterrence.

## B. The Interplay of Different Tort Goals

If market deterrence were the only goal of tort law, then all causation problems would be reduced to a search for the cheapest cost avoider. Tort law, however, is a system that responds to mixed goals. As a result, certain concepts of causation (like causal link) that may be crucial to market deterrence may be ignored if other

[46] When different "type" of damage reflects a situation in which no causal link exists, however, the courts have little trouble in finding no liability. In my gun hypothetical, liability is usually denied because the crucial causal link that I have assumed, that is, loaded guns are particularly hard for infants to hold, is missing. Without it the defendant is almost certainly not the cheapest cost avoider, and collective deterrence or compensation goals do not seem to suffice to impose liability where market deterrence so clearly would not.

[47] That the distinctions are difficult to make is almost tautological. If the distinctions are made in the first place, not because of something inherent in type, extent, or category of plaintiff, but only because these terms provide rough guides for determining relative ability to make a cost-benefit analysis, then the distinctions are bound to be hard to make except on the basis of market deterrence principles.

[48] *See* p. 107 *infra*. The forthcoming book will expand on the notion raised here that clearly focusing on one goal, because it makes that goal too obvious and subject to attack, may actually serve to subvert the goal in a society dedicated to mixed goals.

goals are sufficiently pressing,[49] while other concepts (like proximate cause) may be employed to further these other goals even though they would not come into play if the goals were examined in isolation. If tort law is viewed in this light, it becomes possible to explain certain causation cases which, regarded solely from the standpoint of one goal, are hard to understand.

1. *Collective Deterrence Goals.* I have already noted how the desire to further collective deterrence may modify causation cases. Indeed, *Palsgraf* itself is sometimes explained by saying that the foreseeable wrong was too small to justify liability for the damages that occurred; but this explanation, appropriate enough in collective deterrence terms, conflicts with the doctrine Cardozo enunciated in the case itself. If Cardozo had adopted such a stance he would have promulgated a *Wagon Mound*, rather than a *Palsgraf* doctrine. That he did not do so does not mean that he was insensitive to the pressures of collective deterrence.[50] On the contrary, it simply means that collective deterrence was not his sole aim and that he took into account rough equivalence between wrong and penalty when the influence of other goals either pushed the same way or was inconclusive. Thus, the highly manipulable concept of foreseeability of plaintiff category, which makes some sense in terms of market deterrence but which could be used to respond to collective deterrence pressures as well, might instinctively have seemed the best "doctrinal" answer to Cardozo. When one adds to this Cardozo's desire to center the major part of this difficult melding of goals in the court instead of the jury, the *Palsgraf* doctrine seems quite intelligible.

For reasons I have gone into elsewhere, I do not think it is desirable to mix collective and market deterrence goals in one system.[51] Results like *The Wagon Mound* which hamper market deterrence, but do not go far enough to achieve adequate collective de-

[49] It is possible, though not necessary, to read cases like Kernan v. American Dredging Co., 355 U.S. 426 (1958), and Green-Wheeler Shoe Co. v. Chicago, R.I. & P. Ry., 130 Iowa 123, 106 N.W. 498 (1906), in this way.

[50] Indeed, one way of explaining other Cardozo results is precisely in terms of achieving some equivalence between wrong and penalty. *Compare* Cardozo's opinion in MacPherson v. Buick Motor Co., 217 N.Y. 382, 111 N.E. 1050 (1916)(purchaser allowed to recover when injured by defective automobile), *with* Ultramares Corp. v. Touche, 255 N.Y. 170, 174 N.E. 441 (1931) (creditors not allowed to recover against auditors who had negligently certified the debtor to be solvent), *and* H.R. Moch Co. v. Rensselaer Water Co., 247 N.Y. 160, 159 N.E. 896 (1928) (plaintiffs injured by a widespread fire not allowed to recover against waterworks that negligently failed to maintain necessary water pressure).

[51] *See, e.g.*, G. CALABRESI, *supra* note 2.

terrence, seem all too likely.[52] Yet, so long as our tort system remains a fault system, it must be admitted that the deterrence goals implicit in the system include *both* collective and market deterrence (although, as I have argued, even the fault system emphasizes market deterrence at least as much as collective deterrence). *The Wagon Mound*, though admittedly more internally logical than *Palsgraf*, unlike *Palsgraf*, is clearly wrong because it ignores the fundamental market deterrence aspects of tort law. Cases that ignore altogether equivalency of wrong and penalty (that is, collective deterrence) from my point of view are not similarly wrong; rather they are part and parcel of the movement away from a fault system, which mixes both deterrence goals with compensation aims, toward a non-fault system, which mixes compensation goals with market deterrence alone.

2. *Compensation Goals.* Even more than collective deterrence, compensation aims can explain why causation cases often differ from what they would be in practice if judges and juries responded solely to market deterrence principles. Consider, for example, how spreading and distribution goals influence both the doctrine and the outcome of the cases discussed above concerning unforeseeable (or barely foreseeable) extent of damage, type of damage, and category of plaintiff.

At the doctrinal level, compensation goals will move in the same direction as market deterrence: a party who is the best spreader of foreseeable costs is very likely to be the best spreader of the same type of costs when they simply turn out to be greater than expected. A point may be reached when insurance coverage is exhausted or ability to self-insure no longer exists.[53] But, in the ordinary case, the fact that damages are greater than expected, but of the same sort as expected, would not alter the identity of the party best suited to spread them. Again, the doctrinal (as opposed to the ad hoc) exceptions made by the courts—for example, assumption of risk with respect to a violinist-steel worker's hands—are the very ones in which different knowledge suggests different likelihood of insurance and hence different ability to spread as to the excess.

Where it is the type of damage that is barely foreseeable, on the other hand, the party who can best spread the foreseeable damages

[52] Penalties remain foolishly bound to *but for* relationships. Insurance is allowed when it is not desirable in collective deterrence terms. And the focus on the injurer's limited foreseeability ignores the possible inability of the victim to choose between injury and safety.

[53] It will be instructive to consider specific cases where that seemed likely, to see how the courts deal with such situations. For my definition of "self-insure" see p. 75 *supra*.

is by no means necessarily the best one to spread the unforeseeable loss. Insurance coverage by defendants seems less likely when the unforeseeable damages are of a different type, and thus, even though ability to self-insure would most likely remain unchanged as to such damages, a spreading reason for treating unforeseeable types of damages differently from unforeseeable extent of damages can readily be perceived. It may not, as Robert Morris suggests,[54] be a good reason in practice, but it is at least an understandable one.

Finally, when the category of plaintiff is not foreseeable, the relative likelihood of self-insurance as between defendant and plaintiff, no less than the relative likelihood of insurance, is in doubt. Again, this uncertainty does not mean that an absolute rule, relentlessly applied to exclude defendants from liability to unforeseeable plaintiff categories would be sensible from a spreading point of view. Indeed, unforeseeable plaintiffs may themselves be such poor spreaders that it is more desirable to make the unforeseeing defendant liable. But it is not hard to see how, if it is used only as a flexible starting point, a doctrine that rejects defendant liability to unforeseeable plaintiffs might frequently be consistent with spreading goals.

Doctrines, however, paint in broad strokes; individual cases may present particular spreading problems not dealt with by rules not focused on spreading. Because the proximate cause determination belongs to the jury, proximate cause doctrine is especially well-suited to modifying results in individual fact situations. Since juries do not give reasons, however, it is difficult to gauge the influence of spreading on the decision-making process. Folk wisdom tells us it has great significance. Indeed, fear of the degree to which juries will introduce spreading notions to the detriment of other goals has been given as an explanation for doctrinal rules removing some part of the liability decision from the proximate cause inquiry and hence from the jury as well. If that is indeed the case, one would expect that jury determinations of proximity, when permitted, frequently reflect spreading aims.

More interesting is the fact that even court decisions in particular cases seem to employ proximate cause to achieve spreading goals, at least where market deterrence is not much of a guide. The results in *Gallick v. Baltimore & Ohio R.R.* and *Republic of France & Compagnie Generale Transatlantique v. United States*,[55] which would be hard to explain in terms of ordinary foreseeability of type

[54] Morris, *supra* note 8.

[55] Gallick v. Baltimore & O.R.R., 372 U.S. 108 (1963), was an FELA action by a railroad

and extent of damage rules, can be readily understood if one takes into account the particular spreading situations involved. In *Gallick*, despite a lack of foreseeability in regard to a serious accident caused by insect bites, the railroad was probably the cheapest cost avoider; it was clearly in a better position to pay for such losses now and insure against them in the future. Thus market deterrence and spreading considerations converged to require defendant liability; the Supreme Court so held, brushing aside lack of foreseeability of the type of damages. In the *Republic of France* case (in which the crucial original defendant was the French Shipping Lines), market deterrence might again have leaned slightly toward defendant liability, but spreading leaned the opposite way. The United States, standing in the shoes of victims it had voluntarily compensated in part, sought to shift the whole burden of the loss to the injurer. Scant victim compensation and many other secondary dislocations would have followed. Lack of foreseeability was, realistically, only of *extent* of damages, but the term "proximate cause" proved sufficiently flexible to deny liability without going to a *Wagon Mound* doctrine.[56]

In the *Republic of France* case, spreading goals inconsistent with weak and uncertain market deterrence requirements seemed to dominate; at times, in a system of mixed goals, they may even

employee who was bitten by an insect while working near a stagnant pool on railroad property. The bite became infected, and both of the plaintiff's legs had to be amputated. While the railroad was clearly negligent in maintaining the pool, it argued that there was insufficient causal connection between the negligence and the harm to warrant liability. At trial the jury held for the plaintiff, the state appellate court reversed on the grounds that there had been no causal connection, and the Supreme Court reversed again, reinstating the jury verdict. The Court held that there was sufficient evidence of a causal connection to present a jury question.

Republic of France & Compagnie Generale Transatlantique v. United States, 290 F.2d 395 (5th Cir. 1961), *cert. denied*, 369 U.S. 804 (1962), arose out of the Texas City disaster in which an explosion on the S.S. Grandcamp "on April 16, 1947, resulted in more than 500 deaths, more than 3000 personal injuries, and tremendous destruction of and damage to property." *Id.* at 396. The United States, after winning a series of suits brought against it under the Federal Tort Claims Act, proceeded under the Act of Aug. 12, 1955, ch. 864, 69 Stat. 707-09 (Texas City Relief Act), to compensate uninsured claims up to a maximum of $25,000 per claim. As assignee of partly compensated victims, it pressed claims against the French Lines (the charterer) and the Republic of France (the owner of the ship). One individual claimant, which had been unwilling to assign its claim of nearly $5 million, joined the United States' action. The court, employing proximate cause language, declined to hold the petitioners liable for any claims arising out of the explosion.

[56] Indeed, even in England after *The Wagon Mound*, decisions like Smith v. Leech Brain & Co., [1962] 2 Q.B. 405, in situations similar to *Gallick*, suggest that equivalency of wrong and penalty—the collective deterrence goal of *The Wagon Mound*—may be relaxed if sufficiently in conflict with spreading goals.

dominate fairly clear market deterrence pressures. Occasional cases can be found where liability was imposed even though causal linkage was very doubtful or even nonexistent.[57] Because of the significance of market deterrence such cases are, as one would expect, quite rare. And where they do appear they can frequently be explained as attacks on the concept of fault rather than on market deterrence; causal linkage to wrongful behavior is generally absent, but close causal linkage to other defendant behavior that is not wrongful is present and the defendant would easily be the cheapest cost avoider in a non-fault system of liability. Still, it should not be surprising that, in a system of mixed goals, spreading goals occasionally dictate a result that market deterrence would reject because no relevant causal linkage exists.

Such is the inevitable result of having a system of mixed goals in which discussion of those goals is conducted in an alien langnage—the language of causation. For the principles of causation, while capable of responding to the several tort goals, are not consciously directed at them or at any particular mixture of them. Whether this arrangement is desirable or whether it would be better to eschew causal langnage and to deal with each of the goals self-consciously and explicitly, mixing them according to some, itself self-conscious, formula remains to be discussed.

## III. The Language of Causation and the Goals of Tort Law

Causal requirements, like all other legal requirements, must ultimately justify themselves in functional terms. Law is a human construct designed to accomplish certain goals. Often—perhaps most of the time—the goals are terribly complex and hard to analyze clearly, and one is properly suspicious of analysis and prescription that would discard time-honored legal terms because one cannot find immediate, clear policy justifications for them. Still, the object of law is to serve human needs, and thus legal terms (which in other contexts may have other, deeper meanings) must sooner or later be linked to the service of human needs.

In this sense many seemingly significant philosophical questions concerning cause become irrelevant to the use of that term in law. To amplify: so far as legal language is concerned, the "cause" of a disease would depend on how, at any given time, it could be most easily controlled. From this point of view, in the nineteenth

[57] *See* note 49 *supra*.

century it would have been appropriate to speak of the "cause" of tuberculosis as the absence of sun and the presence of bad living conditions. Other possible factors, that were both *but for* causes and causally linked, were not subject to human control, current or even hypothetical. With the identification of the Koch bacillus all that changed. At first potentially, and subsequently in practice, efforts directed at *this* causally linked element seemed most likely to control the disease. It was unimportant that both living conditions and genetic predisposition also mattered. What use would it be to speak of these elements as causes when living conditions were so hard to modify in comparison to controlling the bacillus and when genetic predisposition was totally outside human control? Yet these two factors remained causally linked to tuberculosis. More recently, the prospect of genetic engineering has again changed the causal language appropriate to this disease. Now one can, in a meaningful way, speak of genetic predisposition as a "cause" of tuberculosis.

The objective, with respect to tuberculosis, is easier to define than are most legal objectives or goals. Causal language in law must respond to the fact that at times the object sought may differ while the factors subject to control remain the same. For example, courts have been able to hold two faulty, independent hunters both liable in cases like *Summers v. Tice*,[58] even though only one of their shots hit the victim and even though a showing of *but for* cause was supposedly a prerequisite for liability. They would not as readily have found two equally faulty independent rapists to be the fathers of an illegitimate child—even though the causal evidence as to the wrongdoing was precisely the same as in *Summers v. Tice* and even though the victim was equally innocent. Rightly or wrongly, the effects (and hence the function) of paternity actions are conceived to be very different from those of money damage claims for injuries. What is "cause" for one need not be "cause" for the other.

If we recognize that *in law* the term "cause" is used in different guises but always to identify those pressure points that are most amenable to the social goals we wish to accomplish, the apparent paradox disappears. Where goals differ, so does the practical definition of causation. Some aspects of causation (like proximate cause) show this functional basis on the surface. Others, like causal linkage and *but for* relationship, often seem to be almost absolute, goal-neutral requirements. But they appear that way simply because they are being employed in contexts where, by and large, the appro-

[58] 33 Cal. 2d 80, 199 P.2d 1 (1948).

priate goals require them to be established almost universally. However, the willingness of courts to ignore not only *but for* causes in cases like *Summers v. Tice*, but even causal linkage where extreme spreading differences seem involved indicates, I suggest, that in the law "cause in fact" (as it was once called), like proximate cause, is in the end a functional concept designed to achieve human goals.

If causal concepts can be used flexibly to identify the pressure points most amenable to our social goals, then use of such concepts has great advantages over explicit identification and separation of the goals. Terms with an historical, common law gloss permit us to consider goals (like spreading) that we do not want to spell out or too obviously assign to judicial institutions. Because, like all moral terms, causal terms have come to have meanings of their own that cannot be changed as a result of one person's analysis, they enable us to resist political pressures that would, if a more "goal conscious," antiseptic language were employed, result in a mixture of goals thought to be less desirable. Finally, and probably most importantly, they enable the introduction of goals we have not been able to spell out or to analyze, but which nonetheless, together with analyzed goals, form part of that set of relationships we call "justice."

If, on the other hand, causal concepts came to be too rigidly defined and applied, they would no longer adequately serve the goals we can analyze nor would they permit the introduction of the goals we cannot affirm too openly or have not been able to analyze at all. Furthermore, such rigidity would mean that the ability to respond to changing goals and mixtures of goals, both analyzed and implicit, which characterizes common law adjudication and concepts, would be lost. No longer could new needs be introduced and old ones dropped without tearing the seamless web. This rigidity has not characterized the way in which cause has traditionally been used in torts, nor is it the way it is being used now in the difficult transition from fault to non-fault systems. As in the past, causal concepts seem to be responding to changes in appropriate pressure points, yet doing so without breaking with the past or with any not clearly understood goals the past contains.

There is, of course, a danger here. Flexible concepts and "unanalyzed goals" can hide what is really at stake, who wins and who loses and why. They may retain the best of a not fully understood past, but they may also serve, plain and simple, to exploit. Clear "radical reform" language eliminates those results that have simply evolved—both useful and exploitative ones. Causal language tends to preserve, for a while, such goals and hence some results that

cannot be easily justified. It is no accident that some commentators will, therefore, be totally impatient with causal concepts that facilitate the retention of unclear goals, while others will find iconoclastic any examination of those concepts that will make them more amenable to goal-oriented analysis.

I am optimistic about our ability to use concepts, like cause, to promote analyzed goals, and, like John Stuart Mill, I am skeptical of our ability to analyze all our goals[59] and, in addition, to acknowledge all that we can analyze. Thus, I am inclined to believe that the requirement of causation in all three of its parts (causal link, *but for* and proximity) will survive in many areas of non-fault liability rather than be replaced by direct appeals to those clearly identified goals which, by and large, those requirements seem to serve. Just because it is flexible and functional, yet responsive to echoes from the past, causal language will survive—that is what will make it as useful in a non-fault world as it was in a fault-based world. Only if it comes to be viewed as "absolute" or as representing inherent "natural" relationships which the law cannot touch, whatever human needs may require, will it outlive its usefulness and be replaced with other concepts in the law.

[59] J.S. MILL, *Essay on Bentham*, in THE PHILOSOPHY OF JOHN STUART MILL 20 (M. Cohen ed. 1961).