Abstract

This chapter begins with a survey of the implicit role of causation in the writings of the early, pathbreaking economic analysts of tort law. It then clarifies the basic distinction between retrospective (ex post) causation and prospective (ex ante) causation, a distinction that forms the core of many subsequent economic discussions of causation. Next, the explicit role of causation doctrines in inducing optimal care and activity levels is examined under the strict liability and the negligence regimes. The analysis is then extended to cover several complications often plaguing the determination of causation: uncertainty over causation, joint actions among tortfeasors and unforeseeability of harm.

JEL classification: K13
Keywords: Foreseeability, Retrospective and Prospective Causation, Optimal Care, Strict Liability, Joint Actions

1. Introduction

The contribution of economic analysis to the clearer understanding of the function of law is particularly evident in the law of causation. The vast juristic literature deliberating the proper meaning of causation has left a trail of doubt and uncertainty. Prominent traditional tort scholars have conceded that ‘there is perhaps nothing in the entire field of law which has called for more disagreement, or upon which the opinions are in such a welter of confusion’ (Keeton et al., 1984, p. 263), and that ‘both courts and textbook writers still fall back when deciding issues in causal terminology’ (Hart and Honoré, 1985, p. 1). Economic analysis helps distinguish the different problems that are involved and offers a unified approach to their resolution.

The attempts of traditional tort scholarship to make sense of the law of causation have led to the classification of the debates into two separate doctrines, cause-in-fact and proximate cause. The cause-in-fact doctrine incorporates the law’s endeavor to define and to conceptualize the criteria that would determine when an act is part of a causal chain that ends with the injury. Here, the but-for test is the most common intuitive criterion for inferring such a factual causality relation. Of all the acts that pass any of the cause-in-fact
tests, the law narrows down the responsibility to those satisfying additional, ‘legal’, tests, which are mostly embodied in the proximate cause doctrine. Liability is imposed only upon a sub-set of the acts that are causally linked to the injury, those that survive the scrutiny of a variety of normative judgments regarding their proximity to the harmful event. As Cooter (1987) nicely labeled it, the proximity doctrine portrays causation as a ‘decaying transitive relation’: as the chain of causal inference extends (‘a caused b, b caused c, ...’), the relationship between removed links weakens without being destroyed.

The economic analysis of the law of causation illuminates both the cause-in-fact and the proximate cause doctrines. Economic analysis applies positive tools from decision theory and statistics to clarify the definition of a cause-in-fact, and to resolve some of the confusion regarding the relative contribution of a given factor to the harmful consequence. Under the normative economic analysis, the proximate cause doctrine’s designated role is to expand or shrink the scope of liability, in order to achieve efficient deterrence.

This chapter is structured as follows: it begins with a survey of the implicit role of causation in the writings of the early, pathbreaking economic analysts of tort law. It then clarifies the basic distinction between retrospective (ex post) causation and prospective (ex ante) causation, a distinction that forms the core of many subsequent economic discussions of causation. Next, the explicit role of causation doctrines in inducing optimal care and activity levels is examined under the strict liability and the negligence regimes. The analysis is then extended to cover several complications often plaguing the determination of causation: uncertainty over causation, joint actions among tortfeasors and unforeseeability of harm.

2. Causation in Early Economic Analysis of Law

The original economic theory of tort law deliberately rejected an explicit role for a causation doctrine in determining liability. Coase’s (1960) view was particularly resolute in its exclusion of a formal causation element. Coase describes an injury as a result of mutual and symmetric interaction among parties. Like particles that randomly collide with each other in space, actions of individuals may conflict and cause one-sided or mutual harm. Thus, the phrase ‘the injurer acted and, when coming across the victim, caused an injury’ is interchangeable with the phrase ‘the victim acted and, when coming across the injurer, caused an injury’. Both passive and active factors are equally necessary in making the harm occur.

Since liability cannot be placed solely on the basis of causation, as both the injurer and the victim are necessary causes, it ought to be decided according to a cost-benefit analysis, which will determine the identity of the party that can alter its actions more cheaply and avoid the injury. As Calabresi (1970)
explained, for instrumental reasons the least-cost avoider should be singled out as the cause of an injury. The forward looking social objective - minimization of accidents’ costs - will be furthered if a party that can prevent an accident with a lower cost than the harm arising from the accident will be regarded as the sole legal cause of the accident and be held liable. Hence, under this view, causation is not a preliminary condition for evaluating liability, but it is the conclusion of the evaluation (see Cooter, 1987).

Landes and Posner (1983, 1987, pp. 228-255) have reinforced this view and argued explicitly that causation has no role in determining liability. Inasmuch as the purpose of tort law is to promote economic efficiency, the injurer should be regarded as the cause of an injury when he is the lower-cost avoider of it, and not otherwise. Therefore, they claimed, ‘the idea of causation can largely be dispensed with in an economic analysis of torts’. When efficiency analysis is conducted to determine liability, it can be fully pursued without reference to causation. Inefficient behavior is synonymous with causing an expected harm.

The symmetry among the roles of the injurer and the victim, as well as the absence of any independent requirement of causation, became well evident when Brown (1973) formulated his rigorous model of accidents. This model - the benchmark for subsequent economic analysis of tort law - assigned symmetric roles to the injurer and the victim, by making the expected harm a function of care levels taken by both. A party’s action can raise the probability of harm and, thus, can only be a cause of an expected harm.

Thus, in early economic analysis of tort law, cause is reduced to efficient prevention: the assignment of legal cause is dependent solely upon the judgment about the economic efficiency of preventive measures. The inquiry into causation carries no additional message once a cost-benefit analysis of the care choices has been completed. This characterization of causation, which prominent scholars have labeled ‘causal minimalism’ (see Hart and Honoré, 1985, pp. lxvii-lxxvii), has led authors to argue that causation serves goals other than efficiency (Epstein, 1973, 1979, 1987; Borgo, 1979; Cooter, 1987) or that it merely represents an older method of conducting efficiency analysis (Grady, 1989; Miceli, 1996).

3. Prospective Causation

Building on the analytical framework of Calabresi (1970) and Brown (1973), subsequent treatments of causation distinguished among two concepts of factual linkage between acts and harms. Calabresi (1975) classified the empirical tests of causality into two types, which he labeled causal link and but-for cause. Both describe effects of actions on outcomes. An act is a but-for cause if, without it, the injury would not have taken place. In contrast, an act has a
causal link to an injury if it increases the probability of its occurrence. As Shavell (1980) later rephrased the distinction, causation can be either retrospective or prospective. Retrospective causation exists if, all else held fixed, but for the action the harmful consequence would not have occurred. Prospective causation exists when an action raises the probability of the harmful consequence. Thus, the distinguishing factor between the two types of causation is the time perspective of the evaluation. Retrospective causation is backward-looking, answering the counterfactual inquiry of whether the action was a necessary condition for the outcome. Prospective causation, in contrast, is forward-looking, answering the ex ante inquiry of whether the action increased the likelihood of injury (see also Rizzo, 1981; Miceli, 1996).

This distinction, and in particular the development of analytical tools to focus on prospective-probabilistic causation, has helped the economic literature advance both in its normative and positive study of the law. In the normative dimension, probabilistic causation became a building block of economic models of tort law. As Shavell (1980, p. 475) has explicitly phrased it, ‘the first-best level of care is determined by the cost of taking care and the degree to which lack of care is a cause of expected losses’. For an action (‘low care’) to raise the probability of a consequence (‘harm’) relative to another action (‘high care’), there must be states of the world in which harm occurs only if that action is taken, and not if the other action is taken.

The prospective causation concept has also advanced the positive analysis of tort law. Perhaps the sharpest example of the contrast between retrospective and prospective causation theories, and the clearest application of prospective causation analysis, arises within the family of ‘coincidental’ accidents cases. In the famous case of Berry v. Borough of Sugar Notch, 43 A. 240 (1899), an excessively speeding streetcar happened to arrive at a point along its route just when a tree fell above that point, and struck it. A strict retrospective causation inquiry would have identified the action of speeding as a but-for cause, since the accident would not have occurred had the streetcar travelled more slowly. Applying the traditional retrospective approach, the court sensed the illogic of assigning liability for such an arbitrary episode, thus had to resort to elusive concepts such as ‘coincidental harm’ or ‘abnormal risk’ in order to screen such results and derive a general principle that will restrict the scope of liability. In contrast, under prospective causation inquiry the action of speeding is recognized to have not affected the likelihood of harm of the type that occurred. Ex ante, a tree can fall at any point along the route, and the speed at which the vehicle is moving does not increase its probability of being hit. The result that the court reached can be easily aligned with the logic of prospective causation. (See Honoré, 1983, pp. 50-55, for early applications of the prospective causation concept.)
4. Causation and Socially Optimal Care

The causation requirement, although not an explicit element in the ordinary economic model of tort law, can be isolated and characterized in economic terms. The basic proposition made by Shavell (1980, 1987, pp. 105-126), and reiterated by Rizzo (1981), Landes and Posner (1983, 1987) and Cooter (1987), claims that the desirability of any precautionary action should be determined only with reference to states of the world in which failure to take the action would lead to greater expected losses. In determining the level of care that is optimal, the benefits of care should be balanced against its costs. But whereas the costs of care accrue before the ensuing state of the world materializes and regardless of the actual state of the world that will materialize, the benefits of care arise only in those states in which taking care reduces harm. For example, in the case of *City of Piqua v. Morris*, 120 N.E. 300 (1918) the defendant failed to take sufficient measures against floods. However, a particularly severe flood occurred, one that even appropriate precautions would not have withstood. Thus, in evaluating the desirability of anti-flood measures, only the chance for moderate floods should be counted.

The idea that care can be cost-justified only with reference to states of the world in which it can reduce the harmful consequence was formulated by Shavell (1987, pp. 118-121) in causation terminology. Shavell defined care (or lack thereof) to be a necessary cause of harm if, given some state of the world, a different level of care would have led to a different level of expected harm. He then proceeded to show that the socially optimal level of care depends only on states of the world in which the injurer’s care would be the necessary cause of any losses that occur.

In order to examine the extent to which liability rules can implement optimal care, a causation restriction was formally introduced to the structure of liability rules. Shavell’s (1980) concept of the scope of liability incorporates the causation restriction. The scope of liability is defined as the set of states of the world under which liability can be applied. The scope of liability is said to be restricted if, given a harmful consequence, there are some states of the world in which the injurer is not held liable. The scope of liability will be unrestricted if, anytime there is a harmful consequence, and no matter what state of the world surrounded it, the injurer will be held liable. The design of a liability regime includes, in addition to the determination of due care (in negligence) and the magnitude of liability (both in strict liability and in negligence), the determination of the scope of liability. If an act is not a necessary cause of the injury, the injury may be left outside the actor’s scope of liability.

Adding the determination of a scope of liability into the analysis nicely extends the early economic models, by capturing the effects of conditions beyond the control of the human actors. When the probability or magnitude of...
injuries depend upon external conditions, analyzing such conditions within the formal structure of liability rules is necessary.

5. Causation Under Strict Liability

Under strict liability, courts have to determine the magnitude of liability and its scope. Assuming the magnitude of liability equals the victim’s actual harm, what remains to be determined is whether the accident is to be included within the scope of liability. Two principal propositions can be made regarding the incentive effects of the determination of the scope of liability:

5.1 The Effect of the Scope of Liability on the Level of Care

The injurer will have optimal incentives to take care as long as the scope of liability includes at least all the states of the world in which the injurer’s care is a necessary cause of the harm. If the scope of liability is too restricted, and does not include all the states in which the injurer could alter the harmful consequence with its care, then the injurer will have insufficient incentives to take care. In this case, the injurer will ignore some of the social benefits of its care - the reduction in expected harm occurring in states of the world outside the scope of liability - and will underinvest in care. If, in contrast, the scope of liability is optimally restricted, and includes only states of the world in which the injurer’s care is a necessary cause, the injurer will bear only the increment in expected losses due to its actions, and will have optimal incentives to take care. Similarly, if the scope of liability is unrestricted, so that whenever harm occurs, and regardless of the state of the world, the injurer is held liable, the injurer will engage in optimal care. Notice that an unrestricted scope of liability does not, in itself, distort the injurer’s incentives to take care. Even if the injurer is liable for harms which its care could not have prevented, it will not exercise excessive care. Taking more care will not prevent the harm in the states of the world in which care is not a necessary cause, and thus will not reduce its expected liability. Hence, the injurer’s incentives to take care can be distorted only by an overly-restricted scope of liability, not by an unrestricted one (see Shavell, 1980, 1987, pp. 105-110; Landes and Posner, 1987, p. 236).

5.2 The Effect of the Scope of Liability on the Level of Activity

If the scope of liability is too restricted, and does not include all the states in which the injurer’s care is a necessary cause, it was already established above that underinvestment in care will arise. This underinvestment can also lead to excessive incentives to engage in the activity, as the injurer will not bear the full ‘externality’ of its activity. The cost of engaging in the activity is reduced by the incremental reduction in the investment in care and by the incremental
reduction in the expected liability and, thus, an injurer may engage in an activity even when it is undesirable from a social point of view. Similarly, if the scope of liability is too broad or unrestricted, it may discourage an injurer from engaging in a socially desirable activity. Although the injurer who faces an unrestricted scope of liability will not take excessive care, the injurer will face an inflated expected liability. As Shavell (1980, 1987, p. 108) has termed it, the injurer may find the unrestricted scope of liability to be ‘crushing’. An activity that is worthwhile may be deterred by imposing upon the actor costs of losses that would have been occasioned regardless of this activity. For example, if a car manufacturer is held liable for accidents arising from bad conditions of roads, such that cannot be avoided by extra prevention devices in the car’s design, it may be led to reduce the volume of production. Hence, for injurers to engage in optimal levels of activity, courts have to restrict the scope of liability appropriately, which, according to some (for example, Burrows, 1984), may demand too much sophistication from the legal system, and, according to others (for example, Wright, 1985) does not reflect prevalent causation doctrines.

6. Causation Under the Negligence Rule

Under the negligence rule, courts have to determine the level of due care, the magnitude of liability and the scope of liability. Assuming that the magnitude of liability equals the victim’s actual harm, what remains to be determined is which harms should be factored into the determination of the standard of due care, and under what states of the world the accident is to be included within the scope of liability. Shavell (1980, 1987, pp. 105-121) has made the following propositions concerning the incentive effects of causal determinations:

6.1 The Determination of the Optimal Standard of Care
The due level of care should equal the optimal level of care, as determined by considering the effect of care only in circumstances in which care is a necessary cause - that is, only in states of the world in which taking care would reduce harm. Care that has no bearing on the occurrence of harm should be excluded from the negligence standard.

6.2 The Effect of the Scope of Liability on the Actual Level of Care
Once a standard of due care is set, the scope of liability has only limited incentive effects. Whether the scope of liability is optimally restricted (to include only states of the world in which the injurer is the necessary cause), or whether the scope of liability is too broad or unrestricted, the injurer will take the due level of care (assumed to be set optimally). Further, unlike the activity-crushing effect of strict liability, under the negligence rule an
unrestricted scope of liability does not necessarily deter the injurer from engaging in the activity. The injurer is induced to take due care and thereby avoid liability, and thus becomes indifferent as to the actual scope of liability (Landes and Posner, 1983, 1987, p. 236). As long as the exaggerated scope of liability does not boost the level of due care, it has no adverse incentive effects per se. In contrast, if the scope of liability is too restricted, and does not include all the states in which the injurer’s care could have reduced harm, the injurer may (but not necessarily) be led to take too little care. The injurer will compare the cost of due care to the cost of liability in its inefficiently restricted scope. If the cost of liability is smaller, the injurer’s incentives to take due care will be distorted.

6.3 The Scope of Liability in an Imperfectly Operating Negligence System

Inasmuch as the application of the negligence rule is plagued with error and uncertainties, it contains an element of strict liability (the injurer may bear liability even if he were not negligent). In this case, the unrestricted scope of liability can have the crushing effect that is associated with the operation of a strict liability rule (Shavell, 1980, 1987, p. 108), Landes and Posner, 1983, 1987, p. 236).

Even after Shavell’s (1980) analysis of the optimal scope of liability, showing that liability should be restricted only to accidents that would not have occurred had the injurer employed due care, most economic models of negligence, including part of Shavell’s (1987) book, continued to implicitly assume that the scope of liability is unrestricted, and that liability turns solely upon the injurer’s negligence. That is, if the injurer were negligent, no matter how slight its deviation from due care, it is liable for any accident that arises, including accidents that additional care would not have prevented. Grady (1983, 1984, 1989), Kahan (1989) and Marks (1994) have shown that the analytical framework which assumes unrestricted scope of liability is not in line with either tort doctrine or optimal incentive design. To adhere to Shavell’s analysis and restrict the scope of liability so that it includes only accidents that were caused by the injurer’s negligence, would imply that an injurer who takes less than due care is not liable for every harm that arises, but only for those harms which would not have arisen had the injurer taken due care. Thus, if the injurer takes slightly less than due care, the proper scope of its liability would include only the (slight) incremental harm that occurred due to this deviation, and will not include all harms that would have occurred anyway. In the ordinary case in which care reduces the probability of an accident but not its magnitude, if the accident occurs the negligent injurer will have to pay damages with a probability less than one.

Using Kahan’s (1989) illustration, suppose the proper height of a fence surrounding a stadium is 10 feet, and the field owner erects a fence of 9 feet. If a ball flies over the fence and causes harm, the scope of liability should be
Causation and Forseeability

(and, as a matter of common law, is) restricted to those accidents caused by balls flying over the fence at heights between 9 and 10 feet. Only those accidents are caused by the field owner’s negligence. Making the field owner liable for all harms caused by flying balls, including those that fly at heights exceeding 10 feet, would mean imposing an unrestricted scope of liability.

Until Kahan (1989) exposed it, most economic models managed to conceal their incorrect characterization of the scope of liability. The reason these models successfully overlooked this restriction relates to the discussion above, which suggested that in the case of a perfectly operating negligence system an exaggerated scope of liability does not have a distorting effect. Since the perfect-information models of negligence find that the injurer will have the proper incentives to take optimal care even under the exaggerated scope of liability regimes, and since there is no crushing of activity effect, they suppress the significance of the exaggerated scope of liability. But, as Kahan clearly demonstrated, an unrestricted scope of liability will have different incentive effects from an optimally restricted scope of liability in cases in which the application of the negligence rule is plagued with information imperfections.

Grady and Kahan’s analyses also suggest that the proper characterization of causation should eliminate what is otherwise considered a prominent feature of the negligence rule: the discontinuity of the injurer’s cost function at the point of due care. This feature of discontinuity plays an important role in models analyzing injurers’ behavior under uncertainty (see, for example, Craswell and Calfee, 1986; Shavell, 1992; Ben-Shahar, 1995). If the injurers’ cost function is continuous, as Grady (1989), Kahan (1989), and Cooter (1989b) have demonstrated it to be, the incentive to deliberately engage in excessive care, to ensure compliance with the uncertain legal standard, is significantly diminished.

7. Uncertainty over Causation

When an injury occurs, its origin may be ambiguous. Several reasons may account for the uncertainty. First, it may be that separate factors created similar risks simultaneously, and the actual injury cannot be clearly traced to any one of them. Second, it may be that the injury manifested itself a long period after the risk was created or the accident occurred, in which case the cause is difficult to identify. The principal question that needs to be addressed in the face of causal uncertainty is under what conditions should a party be liable for injuries that are uncertain to have been caused by its actions?

Two basic approaches to liability in the face of uncertainty over causation can be proposed. The first approach applies an all-or-nothing criterion to determining liability. An all-or-nothing criterion holds that either there is no liability or, if liability is imposed, then it equals the full losses of the victim.
The most common all-or-nothing criterion is the threshold probability rule, under which full liability is imposed upon the defendant if the probability that it caused the accident exceeds a threshold level. Potentially, any threshold can be set, including one that would require proof of causation exceeding any reasonable doubt. However, the prevalent doctrine applying the threshold probability rule is the ‘preponderance of the evidence’ standard of many common law jurisdictions, which incorporates a threshold probability of 50 percent. (In some cases the law reverses the burden of proof and presumes that the defendant is the cause of the injury. Then, the defendant needs to satisfy the 50 percent threshold in proving that he is not the cause of the injury.)

The second approach to resolving uncertainty over causation incorporates a proportional liability criterion. Under this approach, whenever there is a positive probability that the defendant caused the injury, liability will be imposed, but its magnitude will be reduced proportionally to account for the uncertainty. The most common proportional rule, known as the ‘market share’ approach, sets the defendant’s liability equal to the actual harm multiplied by the probability that the defendant caused the injury.

Traditionally, the law of torts has been governed by the first approach of all-or-nothing. Full liability is assigned to a party whose acts are assessed to be a substantial factor in bringing about the harm. A ‘preponderance of probabilities’ - a threshold of 50 percent - is required for imposition of liability, and without it no liability is inflicted. However, beginning in the 1980s, and coming as a response to the onslaught of mass exposure or catastrophic injury torts, American courts have been more willing to apply the second approach. In the case of Sindell v. Abbott Laboratories 607 P.2d 924 (1980), which involved the mass disaster of the DES drug, the court determined each manufacturer to be liable for a fraction of every victim’s harm, with liability determined in proportion to the manufacturer’s market share. The debate over the market share doctrine has since occupied many branches of tort scholarship, including law and economics. The next two sections examine the economic justifications for the two approaches.

8. The Case for Threshold Probability Rules

The first systematic analysis applying economic methods to compare the two liability approaches was presented by Kaye (1982), who proposed to show the superiority of the 50 percent threshold probability rule over any other threshold probability rule as well as over the proportional approach. Kaye’s argument is based on the assumption that in situations of uncertainty over causation the social objective of tort adjudication is to minimize the ex post costs of erroneous liability decisions. Ignoring any ex ante incentive effects that the rules may
have, and assuming that the two types of potential errors courts could make - false positives and false negatives - are just as costly, Kaye shows that the 50 percent threshold rule is the error-minimizing one. To illustrate the essence of Kaye’s argument, consider a case in which the harm is $100 and the probability that the defendant caused it is 40 percent. Under the ‘preponderance of the evidence’ rule, the defendant will not be liable, and the expected error costs will equal $40 (there is a 40 percent chance that the defendant is truly the tortfeasor, in which case it underpays by $100, for an expected error cost of 0.4×$100 = $40). In contrast, if the court applies the proportional liability rule and sets the defendant’s liability at $40, the expected error costs will be $48 (there is a 40 percent chance that the defendant is truly the tortfeasor, in which case it underpays by $60, for an expected error cost of 0.4×$60 = $24; and there is a 60 percent chance that the defendant is not the tortfeasor, in which case it overpays by $40, for an expected error cost of 0.6×$40 = $24; the sum of the expected costs of the two possible errors is $24 + $24 = $48).

The all-or-nothing feature embodied in the threshold liability rule has an additional potential advantage, suggested by Levmore (1990), of reducing the degree of uncertainty. If uncertainty is assumed to be endogenous, and to vary according to the incentives of the parties to bring evidence to court, then the liability approaches can be compared with respect to the degree of uncertainty they generate. Here, Levmore claims, a high threshold probability will produce the most evidence and lead to the least uncertainty. When uncertainty is great and tortfeasors are difficult to identify, plaintiffs face a complete denial of recovery under a threshold rule that sets a sufficiently high threshold probability. This will induce plaintiffs to invest more in developing evidence and identifying the true injurers. In contrast, under a market share regime plaintiffs need not invest in identifying the true injurers, since they are fully compensated regardless of the degree of uncertainty. (Of course, if the defendant, rather than the plaintiff, were assumed to be the party that can develop superior evidence, then a market share rule will give the defendant the greatest incentives to bring evidence.) Again, ignoring the ex ante incentive effects of the rules and focusing on the ex post characteristics of the adjudicatory regime, a case is made for the threshold probability rule.

Apart from minimizing uncertainty and the ex post costs of uncertainty, the threshold probability approach may offer the additional advantage of reducing administrative costs. As Shavell (1987, p. 117) suggests, there are three distinct reasons that the administrative costs will likely be higher under the proportional liability approach. First, the volume of cases is likely to be higher under the proportional liability approach, because actions in which the probability of causation is less than the threshold could be brought. Second, more defendants could be sued in a typical case under the proportional approach, raising the costs of litigating the case. And third, litigation under the
proportional approach requires the added judicial determination of the precise probability of the defendant being the cause of the injury, whereas under the threshold probability approach the only thing that matters is whether the probability of causation exceeds the threshold. Some of these excess costs may be diminished under the enforcement regime that Rosenberg (1984) has proposed, which borrows features from ‘public law’, such as class actions, damage scheduling, and insurance fund judgments.

9. The Case for the Proportional Liability Rule

An economic analysis purporting to show the potential superiority of the proportional rule was presented by Shavell (1985, 1987, pp. 115-118, 123-126). Shavell evaluated the effects of the two approaches with respect to a different social objective from the ex post error-costs minimization criterion. Adopting an ex ante focus, Shavell took the objective of liability regimes to create optimal deterrence, that is, the minimization of accident losses. In this perspective, legal errors do not involve a welfare cost per se, and their minimization could perhaps be taken as a measure of fairness but not as a proxy for optimal deterrence (see also Kaplow, 1994). Once embracing the social objective of optimal deterrence, Shavell has managed to demonstrate that the proportional liability approach is the superior method of inducing socially desirable behavior.

The main propositions made by Shavell are as follows. First, the threshold probability criterion distorts the incentives of parties to take care. If the probability of causation is systematically below the threshold probability, the party will face too little liability and will take less than optimal care. This problem of underdeterrence under the conventional threshold probability rule was labeled by Levmore (1990) and Farber (1990) as the problem of recurring misses. For example, if the probability of a party being the cause of a typical injury is systematically 40 percent (as in the case of a manufacturer holding a 40 percent share of the market for a harm-causing product), the party will always escape liability under the 50 percent threshold rule. The net of liability will miss this party repeatedly. Thus, the party will have no incentives to take care. The underinvestment result arises both under the negligence rule and under strict liability (see also Landes and Posner, 1987, pp. 263-269; Robinson, 1985).

Similarly, if the probability of causation is systematically assessed above the threshold, the injurer may have excessive incentives to take care. This distortion will arise under a strict liability regime, since the injurer will pay for all losses, more than it actually causes. The injurer may take excessive care for a subtle reason. Since the injurer pays for all losses only if he is determined to by the likely cause, the injurer will have the incentive to reduce the chance that this determination would be made. By taking excessive care, the injurer may
be able to reduce the posterior probability that he will be designated as the likely cause of the injury. That is, extra care may shift the preponderance of probabilities, and clear the injurer from liability altogether. Notice that this overinvestment result will not arise under a negligence regime, since the injurer will take due care and will avoid the excessive liability (unless, of course, the level of due care is ill-defined). This is another illustration of the general proposition discussed in Section 6 above, which claimed that an unrestricted scope of liability - that is, liability for consequences that a party did not cause - does not in itself lead to distorted incentives, and only enhances the motive to take due care under a negligence regime.

In contrast to the distorted outcome under the threshold probability rule, the proportional liability approach leads to socially optimal levels of care (see Delgado, 1982; Rosenberg, 1984; Shavell, 1985, 1987, pp. 115-118; Levmore, 1990). The injurer faces expected liability equal to the expected loss associated with its behavior, and will behave as it would in the absence of uncertainty over causation. For example, a manufacturer who causes 40 percent of the harms of a particular type will pay for losses in every case, including the 60 percent of the cases which it did not cause. But in every case its liability will equal 40 percent of the individual harm, thus it ends up bearing liability of 40 percent of the total harm, equal to the fraction it caused.

Another important proposition made by Shavell concerns the incentives of parties to engage in the activity that produces the harm. Again, the threshold probability rule distorts \emph{ex ante} incentives. When the probability of causation is systematically above the threshold, the injurer will be overdeterred from engaging in activity under the strict liability regime. (Under the negligence rule the injurer takes due care and escapes liability, thus engages in the same level of excessive activity as it does in the absence of uncertainty over causation, see Shavell, 1987, pp. 21-32). Likewise, when the probability of causation is systematically below the threshold, the injurer escapes liability and, as a result, engages in excessive levels of activity, both under negligence and under strict liability. In contrast, under the proportional liability rule, the injurer’s incentives to engage in the activity are the same as they would be in the absence of any uncertainty over causation.

Several authors have argued that the market share approach may lead a free-rider problem which will cause underinvestment in care. Marino (1991) demonstrated that care practiced by one firm produces benefits to other firms. By reducing the probability of harm associated with its products, a firm produces a positive externality captured by the other firms in the form of reduced expected liability. That is, each firm will underinvest in care because it bears the cost of care in full but can appropriate only a share of its benefits. The magnitude of this underinvestment will diminish as the firm’s market share increases, and the underinvestment problem will disappear if the firm is
a monopoly. In similar spirit, other authors have argued that proportional liability will not generate optimal incentives for safety research. Delgado (1982) and Rose-Ackerman (1990) have pointed out the public good characteristics of safety improvements, and speculated that the infliction of full liability on an injurer who has the greatest opportunity to conduct safety research may be superior to the division of liability according to proportional causation.

The dichotomy between the proportional liability approach and the all-or-nothing threshold probability approach reflects the tension between ex ante and ex post goals of the tort system. A framework that seeks to unite the two approaches has been offered recently by Porat and Stein (1997). Under this framework, a liability rule should be designed to give incentives to parties who are the cheapest evidence providers, to reduce the ex post uncertainty in assessing liability. The ingenious mechanism these authors examine is titled ‘liability for uncertainty’ - imposing liability for an injury whose cause is uncertain on the party that created or had the best opportunity to prevent that uncertainty. This will lead the parties to invest optimally in removing uncertainties, and when ex post uncertainty is eliminated, the ordinary liability mechanisms can operate to generate optimal incentives to reduce the primary damage. Thus, for example, DES manufacturers can either eliminate the uncertainty over causation (thus avoid liability for uncertainty), in which case the all-or-nothing approach will apply and will generate optimal care incentives, or the manufacturers can choose not to eliminate the uncertainty over causation, in which case they will be liable for the injuries based on their proportional contribution to the creation of uncertainty.

10. Risk-Based Liability and Safety Regulation

The market share approach is a doctrinal step away from the strict fundamentals of causation. But it is not the most radical step. With the growth of mass exposure torts, and due to the large degree of uncertainty over causation in these torts, authors have advocated an even more unorthodox legal mechanism which will practically abandon any causality requirement between the injurer’s action and actual harm. The idea is to structure a liability regime based solely on ‘probabilistic causation’. Under this regime, liability is proportional not to the harm itself, but rather to the risk of harm which the actor produces; it is applied regardless of whether this risk actually materializes. For example, an individual who uses a product and later discovers that she is under a particular risk, that may or may not develop into actual harm, may recover damages equal to her expected harm. Thus, liability is assigned strictly on the basis of the creation of unreasonable risk, independent of any injury. Contrary to the dominant role that the causation requirement was
given in other influential theories (as in Epstein, 1973), here the causation element is essentially eliminated.

Robinson (1985) and Landes and Posner (1984, 1987, pp. 263-269) have argued that awarding its expected losses to each potential victim exposed to the risk of harm will create the proper incentives for injurers to take care and to select the correct activity levels. In the context of mass exposure accidents, and in light of the severity of risk-spreading and bankruptcy concerns, this view has triggered serious attention (see, for example, Roe, 1984; Celli, 1990).

Viewed *ex post*, this approach gives many potential victims a windfall, as they are going to be compensated without actually bearing any risk; at the same time, actual victims will be undercompensated. But viewed *ex ante*, this approach can provide superior incentives for care relative to other approaches that have to await the actual, oftentimes lagged, harm and, thus, dilute the deterrent force of liability. Obviously, a troubling aspect of a risk-based liability regime is its administrative costs. Litigation need not be conditional on occurrence of harm and thus could be more frequent, let alone more complicated (see Celli, 1990). At the same time, each victim is awarded only a fraction of the actual harm, which may reduce the incentives to take legal action and, consequently, will lead to underdeterrence.

A different approach to monitoring incentives in cases that pose inherent difficulties in ascertaining causation is a centralized approach, relying on administrative regulation to enforce optimal risk reduction. Several authors (Shavell, 1984, 1993; Cooter and Ulen, 1989, p. 420) have examined the advantages of relying on regulatory authorities to monitor and deter risk creation in the period before harm manifests itself. These authors have suggested that regulation of safety may perform better than a risk-based liability system in preventing mass torts. The main justifications for the superiority of a regulatory regime are: (1) the government may be a better information-gatherer than the injurer; (2) injurers may be judgment-proof in catastrophic harms, thus liability will not generate sufficient deterrence; (3) suits may not be brought in all cases, due to their costs and to victims’ lack of information, thereby diluting the deterrent effect of liability; (4) administrative costs of decentralized liability regimes may be higher.

11. Causal Apportionment among Joint Tortfeasors

Uncertainty over causation may involve another dimension. Apart from the difficulty of identifying the actual cause-in-fact - the party whose act caused the particular accident - courts may face the difficulty of determining the relative causal *share* of each of several tortfeasors. There may be information as to the probabilistic contribution of each act - what *ex ante* risk each act imposes - and
how the risks change when acts occur simultaneously. However, when two acts operate simultaneously to cause harm, the contribution of each act to the combined risk has to be determined before courts can apply either the threshold probability approach or the proportional liability approach. This determination involves a ‘disentanglement’ of the harm-production process, a logical exercise which has proven to be problematic.

To illustrate the problem, imagine two fires that are set simultaneously and independently and combine to destroy a field. It is estimated that, ex ante, the first fire alone had a 10 percent chance of destroying the field, the second fire alone had a 20 percent chance, and together they had a 50 percent chance of destruction. That is, their joint operation creates a synergistic effect. If both fires were set and the destruction occurred, how should liability be divided across the two ‘causes’? Or, suppose a particular illness can be contracted either by use of a product (1 percent) or, independently, by smoking (4 percent). However, if an individual both uses the product and smokes, the risk increases to 15 percent. Again, a significant synergistic effect exists. What fraction of the harm can a smoker that has used the product recover from the manufacturer? In both examples, how should the synergistic effect arising from the multiple causes joined together be divided across the causal contributors?

The problem of allocating the shares of liability in accidents that have multiple causes is said to have ‘generated a bewildering variety of legal rules and nomenclature’ (Kaye and Aickin, 1984), and have perplexed scholars who expressed ‘doubts that there exists a single factotum satisfactory formula for dividing damages’ (Kruskal, 1986). The first systematic treatment of the causal apportionment problem was offered by Rizzo and Arnold (1980) (see also Rizzo and Arnold, 1986). These authors proposed an apportionment scheme that assigns to each act a share of liability that consists of two elements. The first element is proportional to the act’s ‘marginal product’, which Rizzo and Arnold defined to be the probability of harm given this act operating alone. The second element is a fraction of the synergistic effect, which Rizzo and Arnold arbitrarily selected to be one half.

In the literature that followed, Rizzo and Arnold’s allocation scheme was generally rejected. Kaye and Aickin (1984) found flaws with the statistical framework applied by Rizzo and Arnold, as well as with the material justifications for the apportionment scheme. In particular, Kaye and Aickin suggested that Rizzo and Arnold’s definition of an act’s marginal product - the increase in the probability of harm, given this act operating alone - is no more appropriate than many other possible definitions, such as the incremental increase in the probability of harm, given that the other act is also operating. Kaye and Aickin argued also that there is no one logical way to divide the synergistic effect across the acts, and thus logic alone cannot provide a
Causation and Forseeability

Further criticism of Rizzo and Arnold was voiced by Kruskal (1986), Kelman (1987) and Rose-Ackerman (1990), who again claimed that there is no one principled way to apportion damages that will not be arbitrary and that will make economic sense. An alternative to the Rizzo and Arnold method of defining an act’s marginal product can be derived from the cooperative game-theory concept of the Shapley value. This method offers a more structured way to define an act’s marginal product, based on its expected marginal contribution to the probability of harm, averaged over all possible combinations of acts (see Ben-Shahar, 1996). While this approach enjoys some intuitive appeal that the previous method did not have, it has the similar shortcoming in its reluctance to consider the ex ante effects of the apportionment rule on the incentives for care among multiple tortfeasors.

When multiparties are responsible for an injury, there may not exist an apportionment rule that leads to efficient incentives and keeps total liability equal to harm. To provide the right incentives to all parties, damages exceeding full harm may need to be assigned. For example, Landes and Posner (1980, 1987, pp. 193-201) have focused on the effects of liability apportionment on incentives for care under the negligence regime. They have argued that joint tortfeasors can be led to take due care under a no-contribution rule - that is, a rule that makes each party liable for the entire damage and allows the victim to determine each tortfeasor’s liability share. If the total cost of care is less than the expected harm, at least one of the injurers will have the incentives to take due care (his cost of care is less than half the expected harm), thereby placing the entire liability on the other and leading the other to take due care as well (see also Wittman, 1981, for a related argument in a joint but sequential care setting). Thus, in the celebrated case of Summers v. Tice, 199 P.2d 1 (1948), where two hunters independently and simultaneously fired in the direction of a victim but only one - unidentified - hit, joint liability with no contribution will lead both to take optimal care. Notice that a doctrinal justification for making each hunter fully liable can be obtained through Porat and Stein’s (1997) idea of liability for uncertainty. Each hunter is liable since, but for his action, the apportionment difficulty would not have existed: either the fatal bullet was shot by him, in which case but for his action there would have been no injury, or the fatal bullet was shot by the other, in which case but for his action there would have been no uncertainty.

The problem with the negligence-based no contribution rule is that it may lead injurers to engage in excessive levels of activity. Miceli and Segerson (1991) study a different apportionment rule, one that would potentially lead joint tortfeasors to take efficient care and make efficient activity decisions. They propose a ‘decoupled’ strict liability regime, under which each tortfeasor pays

a sum equal to the difference between actual damages and the damages that would have resulted were he inactive. That is, all tortfeasors are held strictly liable simultaneously, but each receives a ‘credit’ for the expected damages that would have occurred in his absence. Since this rule may lead to payment exceeding actual harm, the excess can be paid as a fine to the state.

12. Unforeseeability

Whether the objective probability of an accident is low or high should not in itself affect its inclusion or exclusion from the scope of liability. As Shavell (1980) explained, even if the probability of the harmful consequence is very small, the act that is the cause of the increase in probability should carry liability. The effects on incentives to take care and on the level of activity will be correspondingly small, as they ought to be. In addition, the added expected administrative costs of adjudicating a low-probability event are small, since these costs will be incurred only with a small probability. Thus, the benchmark claim is that whenever an act is a necessary cause of the harm, liability should follow (see also Rizzo, 1981).

However, to determine the incentive effects that any scope of liability generates, it is not the objective probability of harm that matters, but the subjective probability - the ex ante assessment of the possible consequences as it is made by the injurer. Calabresi (1975) was the first to state explicitly that there is no sense in trying to deter an act which is a necessary cause of the injury by threatening to impose liability on an injurer who assigns a subjective probability of zero to the injury. An injurer who does not foresee a harmful consequence cannot be meaningfully labeled the least-cost avoider. As Shavell (1980) clarified and generalized, whenever the subjective probability is very low in absolute terms, and lower than the objective probability in relative terms, liability will not produce sufficient ex ante behavioral effects to justify the increase in the costs of dispute resolutions. Thus, under the doctrine of unforeseeability, accidents whose probabilities are likely to be underestimated by injurers should be excluded from the scope of liability. (But see Rizzo, 1981, for an alternative view, advocating the use of objective probabilities in determining the abnormality of events.)

It may be that an injurer does not foresee some specific low-probability consequences that subsequently materialize. However, the same injurer may still be in a position to associate an activity with unforeseen risks. The injurer may recognize a large variance of outcomes even if it does not recognize the nature of each outcome in the distribution. In this case, assigning liability for unforeseen harms will have the desirable effect of reducing the level of an activity that is known, ex ante, to cluster many unforeseen risks (see Shavell,
Causation and Forseeability

For example, handing a loaded gun to a child leads to many unforeseeable risks (apart from the obvious ones), and can be deterred by imposing liability on this action. Of course, when injurers systematically fail to recognize the unforeseeable consequences of their actions, other forms of deterrence may be required, such as criminalizing the actions (see Calabresi, 1975).

The scope of liability for low probability events has another important effect, in determining the incentives of potential injurers to investigate and contemplate the potential consequences of their actions. That is, the amount of information individuals have regarding risk and risk avoidance can be thought of as endogenous, influenced in part by liability rules. The effects of liability rules on the incentives to acquire accurate information \textit{ex ante} have been studied in more general contexts by several authors (see, for example, Kaplow and Shavell, 1992, 1996; Shavell; 1992; Ben-Shahar, 1995). Specifically, as the scope of liability for low-probability harms expands, individuals will have greater incentives to learn and anticipate these harms, and to take proper measures to avoid them (or liability for them). Thus, the unforeseeability doctrine should be replaced by a doctrine of ‘expensive foreseeability’: only risks that are too costly to anticipate and foresee will be excluded from the scope of liability. Therefore, in operating the hand formula to determine whether lack of care should be considered negligent, courts have to account not only for the direct costs of care, but also for the costs of foresight (Calabresi, 1975; Grady, 1984; Landes and Posner, 1987, pp. 239-247).

The concept of unforeseeability in tort law differs from the concept of unforeseeable damage in contract law. The tort doctrine excludes only ‘freak’ accidents from the scope of liability and serves to avoid unnecessary dispute resolution costs. The contract doctrine serves only to limit damages to the average, reasonable, level. However, in economic terms, both tort and contract doctrines can be justified by their effects on \textit{ex ante} information acquisition. The limited scope of liability in tort should apply only when the risk is too costly for the injurer to anticipate, thus it creates incentives to foresee. The limited scope of liability in contract has the important effect of inducing high-risk contractual parties to reveal their idiosyncratic traits \textit{ex ante}, so as to secure full compensation and to reduce the chance of breach (see Bebchuk and Shavell, 1991).

Two prominent tort doctrines can further illustrate the role of foreseeability in monitoring incentives. The first doctrine distinguishes between \textit{categories} of harms and can be illustrated by the well-known case of \textit{Palsgraf v. Long Island R.R.} (162 N.E. 99 (1928)). In that case, as the result of a railroad employee’s negligence, a parcel containing fireworks fell from the train, exploded and caused the crowd to panic and to knock down scales that were standing on the other platform, in a manner that injured a passenger. Since the
employees did not know of the parcel’s content, the court found the harm to be unforeseeable and outside the scope of negligence. This result is found by most economic writers to be justified (see Calabresi, 1975; Shavell, 1980; Landes and Posner, 1983, 1987, pp. 246-247). Since the injurer discounted the probability of that type of accident, liability would not have generated better incentives and would not have led to prevention of the accident. The injurer may be the least-cost avoider with respect to losses from injuries to passengers and, at the same time, not the least-cost avoider with respect to non-passengers. Making the railroad liable for injuries to non-passengers from acts that were negligent due to the risk they posed to passengers will either have no effect in reducing risks to non-passengers, or may lead the railroad to engage in socially excessive investments to identify freak accidents.

Another well-known tort law doctrine is the eggshell skull doctrine, according to which courts impose liability for bodily harm equal to the full severity of the injury, even if the extent of the injury was unforeseeable due to a pre-existing condition of the victim. This may seem to contradict the basic economic insight, which established that when the probability of a high loss is systematically underestimated, holding the injurer liable for the total loss does not increase the incentive to take care (Shavell, 1980). However, as Landes and Posner (1983, 1987, pp. 249-250) have explained, there is economic sense in holding liability for an unforeseeable extent of injury. First, allowing only average damages could potentially generate optimal care incentives, but only if carried out in all cases, including in cases in which victims have ‘thick skulls’ (that is, awarding these victims average damages despite the fact such damages are known to overcompensate them). Put differently, if at the high end of the distribution of harms liability is capped, but at the low end of the distribution liability equals actual harm, average liability will fall short of average harm and the incentives for care will be diluted. As long as victims with low harms get their actual damages, victims with eggshell skulls should also get their full damages, to maintain the correct level of average liability \textit{ex ante} (see also Kaplow, 1994; Kaplow and Shavell, 1996). In addition, liability for an unlimited extent of injury may have a desirable activity-reducing effect, resulting from the injurer’s de facto strict liability.

13. Causation and Litigation Costs

The restriction on the scope of liability that the causation requirement embodies has, in itself, an ambiguous effect on the administrative costs of the legal system. A restricted scope leads to a lower volume of suits that are filed, saving the litigation costs in cases that are clearly outside the scope of liability. On the other hand, if the scope is unrestricted there may be less harms (through a
Causation and Forseeability

reduction in the level of activity) and thus less suits, and each suit that is filed may be less costly to litigate, as there is no causation issue to resolve (see Shavell, 1980, 1987, p. 109; Landes and Posner, 1983).

Informational imperfections and their legal treatment have important effects on the costs of resolving disputes. When courts are uncertain about causation, a significant portion of the trial effort may be devoted to disentangling the causation process. Applying simple standards such as the 'preponderance of the evidence' rule may reduce administrative costs, sufficiently so to overshadow its inferior incentive effects. Similarly, when the court’s ex post assessment of the probability of harm exceeds the injurer’s ex ante assessment, administrative costs of determining liability may tip the scale towards categorizing the harm as unforeseeable. Lastly, when causation is difficult to verify ex post, but probabilistic linkage is known to exist ex ante, the costs of decentralized dispute resolutions may exceed the costs of a centralized regulatory scheme, outlining the proper bounds of the civil liability system.

Bibliography on Causation and Foreseeability (3300)

Ben-Shahar, Omri (1996), 'Multiple Causation and the Shapley Value: A Note on Apportionment of Damages', Unpublished Manuscript, Tel-Aviv University.
Kahan, Marcel (1989), 'Causation and Incentives to Take Care under the Negligence Rule', 18 Journal of Legal Studies, 427-447.


Rossello, Carlo (1990), Il danno Evitabile. La Misura della Responsabilità tra Diligenza ed Efficienza (Avoidable Damage: The Measure of Liability Between Due Care and Efficiency), Padova, CEDAM, 320 p.


**Other References**

Keeton, W. Page, Dobb, Dan, Keeton, Robert and Owen, David (1984), Prosser and Keeton on Torts (5th edn), St Paul, MN, West Publishers.
Cases

Berry v. Borough of Sugar Notch, 43 A. 240 (1899).
City of Piqua v. Morris, 120 N.e. 300 (1918).
Palsgraf v. Long Island R.r. 162 N.e. 99 (1928).