

- Green, E.J. and Porter, R.H. 1984. Noncooperative collusion under imperfect price information. *Econometrica* 52: 87–100.
- Griffin, J.M. 1989. Previous cartel experience: any lessons for OPEC? In *Economics in Theory and Practice: An Eclectic Approach*, ed L.R. Klein and J. Marquez, Boston: Kluwer.
- Jacquemin, A., Nambu, T. and Dewez, I. 1981. A dynamic analysis of export cartels: the Japanese case. *Economic Journal* 91: 685–96.
- Jacquemin, A. and Slade, M.E. 1989. Cartels, collusion, and horizontal merger. In *Handbook of Industrial Organization*, Vol. 1, ed. R. Schmalensee and R. Willig, Amsterdam: North-Holland.
- Lean, D., Ogur, J. and Rogers, R. 1985. Does collusion pay . . . does antitrust work? *Southern Economic Journal* 51: 828–41.
- Palmer, J. 1972. Some economic conditions conducive to collusion. *Journal of Economic Issues* 6: 29–38.
- Phlips, L. 1995. *Competition Policy: A Game-Theoretic Perspective*. Cambridge: Cambridge University Press.
- Porter, R.H. 1983. A study of cartel stability: the Joint Executive Committee, 1880–1886. *Bell Journal of Economics* 14: 301–14.
- Posner, R. 1970. A statistical study of antitrust enforcement. *Journal of Law and Economics* 13: 365–419.
- Selten, R. 1973. A simple model of imperfect competition, where 4 are few and 6 are many. *International Journal of Game Theory* 2: 141–201.
- Shapiro, C. 1989. Theories of oligopoly behavior. In *Handbook of Industrial Organization*, Vol. 1, ed. R. Schmalensee and R. Willig, Amsterdam: North-Holland.
- Smith, A. 1776. *An Inquiry Into the Nature and Causes of the Wealth of Nations*. New York: Modern Library, 1937.
- Stigler, G.J. 1964. A theory of oligopoly. *Journal of Political Economy* 72: 44–61.
- Suslow, V.Y. 1988. Stability in international cartels: an empirical study. Hoover Institution Working Paper E-88-7, Stanford, CA.
- Telser, L.G. 1980. A theory of self-enforcing agreements. *Journal of Business* 53: 27–44.
- Werden, G.J. and Baumann, M.G. 1986. A simple model of imperfect competition in which four are few but three are not. *Journal of Industrial Economics* 34: 331–5.

causation and tort liability. We say that a person's act caused harm if the harm would not have occurred had the person not committed the act. More generally, we say that X is a cause of Y if Y would not have occurred in the absence of X. This is the principal meaning of causation and the one which will usually be employed here. It is sometimes referred to as causation in fact, 'but for' causation (as in 'but for X, Y would not have occurred'), or necessary causation, to distinguish it from other concepts of causation that fall under the heading of 'proximate causation'. Proximate causation will be addressed in a separate section below.

A fundamental characteristic of tort liability law is that a party must have caused harm in a relevant sense to be held liable for it. The main question to be examined here is how this feature of tort liability affects deterrence and the administrative costs of use of the legal system (compensatory goals will be commented upon in conclusion). What will be said about this question in the tort context will have clear relevance also to civil law generally and to criminal law.

CAUSATION AND STRICT LIABILITY. The primary advantage of holding parties strictly liable for harm if and only if they caused the harm is that this creates socially desirable incen-

tives for parties to take precautions to reduce harm and to engage appropriately in activities that may cause harm.

Suppose that a firm's production generates pollution and that the pollution can cause house paint to peel; the associated cost of repainting homes is \$100,000. Suppose too that peeling may also come about from other factors (prolonged exposure to the sun) and that the cost of repainting homes for this reason is \$80,000. It is then socially desirable for the firm to invest in a device like a smoke scrubber to eliminate the pollution if and only if the device costs less than \$100,000, for that measures the increase in social harm due to the pollution. Further, this is precisely how the firm will be motivated to act if it is liable for repainting costs if and only if it is the cause of peeling house paint. In particular, if the firm is liable even when it is not the cause of losses, it will have an excessive incentive to spend on precautions. In that case, as its pollution-associated liability would be \$180,000, the firm would be willing to spend up to \$180,000 on the smoke scrubber to eliminate pollution and avoid liability for the costs of repainting homes.

It might be that the firm cannot avert the harm by taking precautions. In this case, the firm can, of course, discontinue the activity that generates pollution. Clearly, it should do so only if the benefit from the activity is less than \$100,000; it is socially desirable for the firm to continue with its activity if the benefit from the activity exceeds \$100,000. Again, this is what the firm will do if it is liable for repainting costs if and only if it is their cause. If it is liable for all repainting, the firm may be undesirably discouraged from continuing with its activity.

The basic function of the causation requirement under strict liability, in other words, is that it furnishes socially appropriate incentives to reduce the risk of harm by imposing liability equal to the increase in social costs due to a party's actions. This tolling-of-social-costs function of the causal requirement was initially emphasized in an important article by Calabresi (1975) and was amplified by Shavell (1980) in the first formal economic treatment of causation and liability.

CAUSATION AND THE NEGLIGENCE RULE. The advantage of holding parties liable for negligence if and only if their negligent behaviour caused harm also relates to incentives to reduce risk, but the logic supporting this statement is somewhat different from that applying under strict liability, and other factors are involved as well. Suppose that it would be negligent for the owner of a building to fail to put treads on the stairs in the building. If the owner does not install treads and a person injures himself on the stairs, the injury might or might not be caused by the absence of treads. If the injury is so caused, the owner will be liable. If, however, an injury that occurs on the stairs is not caused by the absence of treads – perhaps a person trips on an obstacle – the owner will not be liable for failure to install treads. Therefore, the principle of no liability in the absence of causation dilutes the expected sanction for not installing treads on the stairs.

Will the threat of liability for negligence when but only when negligence is the cause of losses result in a proper incentive to be non-negligent? Specifically, could the fact

that a party will sometimes escape liability for negligence due to the causal requirement result in too little incentive to take precautions? The answer is no. Rational actors will always be led to act non-negligently, assuming that the standard of due care determining what constitutes negligence is correctly calculated by courts; see Shavell (1980, 1987:105–108). The proof of this conclusion involves, among other elements, the point that the socially desirable level of care itself implicitly reflects causation; care is socially valuable only to the degree that it can reduce accident losses in circumstances where losses would otherwise result.

Although allowing parties to escape liability if their negligence does not cause losses still leaves sufficient deterrence to induce proper care-taking, an affirmative reason for insisting on causation before imposing liability has not been supplied. What would be the disadvantage of imposing liability for negligent acts even when these acts did not cause harm? At first consideration, it seems that there would not be any problem of over-deterrence (as there is under strict liability). Under the negligence rule, actors would just have a stronger reason to act non-negligently. To the extent, however, that there are errors in the negligence determination, the negligence system takes on aspects of strict liability and there could thus be an issue of over-deterrence.

Moreover, there exists a potential administrative cost advantage of imposing liability for negligence only when there is causation: this reduces the scope of liability and therefore may reduce the administrative costs of the legal system because fewer cases are brought. Nevertheless, in each case that is brought issues of causation may need to be resolved (whether or not the absence of treads or something else caused an accident may need to be ascertained). Consequently, it is possible that use of the causal requirement for liability increases rather than reduces administrative costs.

A different issue concerns the often-emphasized point that under the negligence rule a person's expected liability rises discontinuously with his level of care: liability is zero if a person is not negligent yet becomes distinctly positive as soon as his behaviour crosses the negligence threshold. For example, if it is negligent to drive at speeds exceeding 50 mph, a person's liability will be zero if he drives at speeds up to 50 mph but will rise suddenly if he drives at 51 mph, for then he will be liable for all accidents that he causes. This jump in expected liability makes the incentive to be non-negligent sharp, which has both socially advantageous and disadvantageous aspects. It may be socially advantageous because it means that parties will have incentives to be non-negligent even if they cannot pay for the entire harm, or even if they will not always be sued for harm (Cooter 1982; Shavell 1987: 167–8). And it may be socially disadvantageous because it means that parties may be led to take excessive care to reduce the risk of mistakenly being found negligent and bearing liability; if people drive at 45 mph they will lower the chance of erroneously being clocked at a speed of over 50 mph (Craswell and Calfee 1986).

But, as originally noted by Grady (1983) and Kahan (1989), there may not be a sudden increment in expected

liability – expected liability will rise continuously – if liability for negligence is properly limited by the causal requirement under discussion. If a person drives at 51 mph, it might be thought that he should be liable only for accidents that were caused by going the extra mile per hour beyond 50 mph: the driver should not be liable for accidents that would have occurred had he been traveling at 50 mph or less. Hence, the possible social advantages and disadvantages associated with the discontinuity in expected liability may be rendered irrelevant by the causal requirement for liability. However, if the court is unable to tell whether or not an accident would have occurred had a person been driving more slowly, the person driving 51 mph will be liable for any accident that his driving causes, and there thus would be a jump in his expected liability.

UNCERTAINTY OVER CAUSATION. In many situations there is uncertainty about causation. For example, it may not be known which manufacturer out of many sold the product (a drug, lead paint) that caused injury, or whether an injury was caused by the defendant or background factors (was cancer caused by a firm's pollutant or by unknown environmental or genetic determinants?).

The law takes two approaches in such situations. The traditional approach is to hold a defendant liable if and only if the probability that the defendant was the cause of losses exceeds 50%. This approach may lead either to inadequate or to excessive incentives to reduce risk. Suppose that a firm sells to only 20% of the market. Then the likelihood of the firm being the cause of losses from a product-related injury will lie below the 50% threshold and it will escape liability for any harm caused by its product. Consequently, the firm will have no liability-related incentive to take precautions. If, however, a firm's market share exceeds 50%, the firm will be liable for all harms due to the product it sells – and for all harms due to the products that others sell – for it will always be correctly said to be more likely than not the cause of harm. Thus, the firm's liability burden will be socially excessive. These potential problems of inadequate and of excessive incentives may arise under any liability criterion based on a threshold probability of causation; they are not unique to a 50% threshold. Essentially this point has been frequently mentioned (see, for example, Tribe 1971, and Landes and Posner 1983) and is formally developed in Shavell (1985).

The second approach that the legal system has taken is to hold defendants liable despite any uncertainty over causation but to impose damages only in proportion to the likelihood of causation. Thus, liability has been imposed according to the share that firms have in the market for a product, according to the decision in *Sindell v. Abbott Laboratories*, 163. (However, the proportional liability principle has so far not been extended to situations where the alternative causes are background factors rather than other defendants selling the same product.) Under the proportional liability principle, it is readily shown that incentives to reduce risk are proper. If, for example, a firm has 20% of the market, it will pay 20% of harm in every case, so that its liability bill will be the same as if it pays for all the harm in the 20% of cases it truly causes – in which case we know that safety incentives will be socially appro-

priate. That the proportional liability principle engenders optimal incentives (without there being a need to establish causation in particular cases) is an advantage of the principle relative to the traditional threshold probability criterion; see Rosenberg (1984) and Shavell (1985, 1987:115–18).

Yet a disadvantage of the proportional approach is that it could lead to a substantial increase in the volume, complexity, and thus cost of litigation, for under the proportional approach any party for whom the probability of having caused a loss is positive can be sued and have to pay damages.

PROXIMATE CAUSATION. Even if a party is shown to be a cause of losses, he may still escape liability because he was not the proximate cause of losses, where this term has two major meanings (additional meanings will not be reviewed here). One connotation of proximate cause is that harm came about in a 'direct' or expected way, rather than in an unusual, freakish manner. The latter is illustrated by a famous case (*Palsgraf v. Long Island R.R.*, 248) in which, because a man was pushed on board a train just as it started to move, he dropped a parcel; the parcel turned out to contain fireworks; these fell under the wheels of the train and caused an explosion, the concussive force of which was said to have toppled a scale that struck and injured a woman who was sitting on a bench some distance away. Allowing parties to escape liability for very unusual accidents is sometimes thought not to undermine deterrence on the ground that no one could have foreseen such accidents. This argument, however, is subject to the criticism that courts may find it difficult to discriminate between accidents that can and cannot be foreseen. Moreover, the argument may lead to the *reductio ad absurdum* that liability should never be found: any accident, after all, can be seen as extremely unlikely if it is described in sufficient detail. On incentives and liability for unlikely accidents, see Shavell (1980, 1987:128–30).

A second notion of proximate causation is illustrated by two cases in which proximate cause is said to be lacking. In one case (*Berry v. Sugar Notch Borough*, 191) a speeding bus happened to be at just the 'right' point to be struck by a falling tree. Here, note that although the excessive speed of the bus did cause the accident, the accident would have been avoided if the bus had been going faster just as much as if it had been travelling more slowly. In the other case, a person negligently handed a loaded gun to a child to be used as a plaything and the child dropped the gun on his toe, suffering an injury. In such cases, liability is not found because of lack of proximate cause in the sense that the accidents are coincidental to defendants' behaviour, unrelated to the normal risk created by their behaviour.

It can be demonstrated that allowing parties to escape liability for accidents like these does not lead to inadequate precautions. Holding a bus company liable when trees fall down on buses will not induce the company to have its buses go more slowly, for one presumes that the probability of a bus being struck by a falling tree does not depend on the speed of the bus. Likewise, holding a person liable when an object he gives to a child drops on the child's toe will not induce people to remove bullets from guns, for the

probability of a gun dropping on the child's toe will not be affected by its being loaded (setting aside the negligible weight of the bullets). See Calabresi (1975) and Shavell (1980, 1987:110–15).

An advantage of permitting defendants to escape liability when accidents are coincidentally caused is that this will lower legal administrative costs by reducing the scope of liability, unless the cost of deciding about the issue of coincidental causation exceeds the savings from the reduction in the scope of liability. A disadvantage of allowing defendants to escape liability for coincidental accidents is that this means that actors do not bear the full increase in social costs due to their activity (if people did not ride in buses, they might not be struck by falling trees); and the control of levels of activity is an object of, at least, strict liability.

CONCLUDING COMMENTS. In the foregoing discussion, the focus was on the relevance of the legal treatment of causation to incentives and to administrative costs, but not to compensation of victims. Notably, that victims do not receive compensation when injurers escape liability on causal grounds was not treated as socially disadvantageous. However, victims' desire for compensation will generally not depend on the cause of losses, only on the magnitude of losses. Accordingly, the causal requirements of tort law do interfere with satisfaction of the goal of compensation of victims against loss. But this consideration is limited in importance to the degree – which is substantial – that private or social insurance can provide compensation to victims (indeed at lesser cost than the legal system).

Finally, it should be remarked that most commentators and jurists conceive of causal requirements for imposition of liability as arising from the deeply-rooted notion that it is unfair to punish a person for harm unless we can confidently say that that person, instead of someone or something else, caused the harm. Viewing causal requirements in this way, rather than as advancing instrumental social ends, is not intellectually satisfying and can lead to undesirable legal practices (such as failure to adopt proportional liability when that is needed to create proper incentives to reduce risk).

STEVEN SHAVELL

See also DUE CARE; LAST CLEAR CHANCE; LEGAL STANDARDS OF CARE.

Subject classification: 5d(ii).

CASES

- Berry v. Sugar Notch Borough*, 43 A 240 (Pa 1899).
Palsgraf v. Long Island Railroad Co., 162 NE 99 (NY 1928).
Sindell v. Abbott Laboratories, 607 P2d 924 (Cal. 1980).

BIBLIOGRAPHY

- Calabresi, G. 1975. Concerning cause and the law of torts. *University of Chicago Law Review* 43: 69–108.
 Cooter, R. 1982. Economic analysis of punitive damages. *University of Southern California Law Review* 56: 79–100.
 Craswell, R. and Calfee, J. 1986. Deterrence and uncertain legal standards. *Journal of Law, Economics, and Organization* 2: 279–303.

- Grady, M. 1983. A new positive economic theory of negligence. *Yale Law Journal* 92: 799-829.
- Grady, M. 1984. Proximate cause and the law of negligence. *Iowa Law Review* 69: 363-449.
- Kahan, M. 1989. Causation and incentives to take care under the negligence rule. *Journal of Legal Studies* 18: 427-47.
- Landes, W. and Posner, R. 1983. Causation in tort law: an economic approach. *Journal of Legal Studies* 12: 109-34.
- Rosenberg, D. 1984. The causal connection in mass exposure cases: 'public law' vision of the tort system. *Harvard Law Review* 97: 849-929.
- Shavell, S. 1980. An analysis of causation and the scope of liability in the law of torts. *Journal of Legal Studies* 9: 463-516.
- Shavell, S. 1985. Uncertainty over causation and the determination of civil liability. *Journal of Law and Economics* 28: 587-609.
- Shavell, S. 1987. *Economic Analysis of Accident Law*. Cambridge, MA: Harvard University Press.
- Tribe, L.H. 1971. Trial by mathematics: precision and ritual in the legal process. *Harvard Law Review* 84: 1329-93.

centralized and decentralized regulation in the European Union. Raising and spending tax revenue are not the only important ways in which the state controls economic activity: different organs of the state also regulate the activities of private sector agents in a variety of ways. In a society of any size, important questions arise about the appropriate degree of centralization of these regulatory powers, and consequently about the comparative advantage of central and local jurisdictions in carrying out particular regulatory functions. These questions have been prominent in the recent politics of the European Union (see van Keersbergen and Verbeek 1994), particularly since the holding of referenda by a number of member states to ratify the Treaty of Maastricht (1991). Whereas in many other societies issues about the right degree of centralization have been posed in the context of pressures for self-determination or even secession by geographical or ethnic minorities, in the European Union they have arisen chiefly as an accompaniment to a gradual centralization of regulatory powers from member states to the institutions of the Union itself.

Article 3b of the Treaty of Maastricht introduced explicitly the principle of subsidiarity as a determinant of the allocation of powers between the Union and its member states. The relevant clause is worded as follows:

In areas which do not fall within its exclusive competence, the Community shall take action only if and in so far as the objectives of the proposed action cannot be sufficiently achieved by the Member States and can therefore, by reason of the scale or effects of the proposed action, be better achieved by the Community.

Though there is substantial ambiguity about almost every element of this clause (see Begg et al. 1993: 14-23), its spirit is reasonably plain: when in doubt, action by member states is to be preferred to action by the Community, and the burden of proof should lie on those proposing Community action. This principle provokes several fundamental questions. First, is it indeed reasonable to prefer decentralized allocations of power to centralized allocations in the absence of contrary arguments? Secondly, what might count as good contrary arguments, as good reasons

for centralization in specific contexts? Thirdly, to what extent does the actual allocation of regulatory powers in the Community conform to the subsidiarity principle?

So what are advantages of decentralization? Systematic debates about the right degree of centralization of the powers of the state go back at least to the mid-eighteenth century, and were central to the process that led to the drawing-up of the American Constitution. Montesquieu's arguments in *De L'Esprit des Lois* (1748) for the separation of powers as a guard against tyranny were influential on the deliberations of the American founding fathers. They understood separation to mean both the functional division of responsibility between different branches of government (such as legislature, executive and judiciary), and the decentralized retention of power by states as opposed to the federal government. In the nineteenth century de Tocqueville (1835) saw decentralization as one of the great virtues of American political and social life.

However, formal economic analysis of decentralized government has been remarkably limited to date (in contrast to the enormous attention devoted to decentralized markets), and has not even reached a secure consensus as to its main strengths and weaknesses. There are two main traditions in the literature on public finance. One, represented by Oates (1969), supposes that there is spatial variation in the preferences of citizens for public goods (which may in turn be due partly to variations in local conditions). Decentralization allows different kinds and levels of public goods to be supplied in different localities. Centralized governments are supposed to be unable to practise such differentiation, either because they lack information about local preferences and conditions, or because they are constrained to make uniform provision for some other reason. However, they are able to exploit scale economies (important for such public goods as national defence), and to internalize externalities between localities (as when national highways benefit traffic between as well as within regions). For each type of public good, therefore, there will be some level of government that optimally balances these advantages and disadvantages.

A second tradition, due to Tiebout (1956), treats the spatial variation in preferences for public goods as partly the result rather than as purely the cause of the local differentiation of public good supply. If citizens are costlessly mobile between jurisdictions they can choose to locate in jurisdictions that offer their preferred combinations of local public goods, financed by lump-sum local taxes which can be thought of as the 'price' of purchasing a given combination. In effect local governments are like firms offering differentiated products, and Tiebout showed that such provision would be Pareto-efficient provided certain (stringent) conditions were met, notably that the number of jurisdictions was at least equal to the number of types of consumer and that there were no externalities between jurisdictions (see also Pesticau 1977 and Bewley 1981).

Neither of these two traditions answers the question why decentralization of power to local governments should be at all necessary for the local differentiation of public good supply. The Oates model simply assumes that central governments have no access to local preferences, whereas there is no reason why they could not - if they wished -