THE MODERN LAW REVIEW

Justifying Exceptions to Proof of Causation in Tort Law

Sandy Steel*

This article defends a set of exceptions to the general rule in tort law that a claimant must prove that a particular defendant's wrongful conduct was a cause of its injury on the balance of probabilities in order to be entitled to compensatory damages in respect of that injury. The basic rationale for each exception is that it provides a means of enforcing the defendant's secondary moral duty to its victim. The article further demonstrates that the acceptance of this set of exceptions does not undermine the general rule.

INTRODUCTION

On most moral views, causation matters to individual liability to pay compensation for another's injury.[1] Causation provides both a reason for and a limit to the imposition of such liability.[2] On these views, only individuals who bear a causative relationship to another's injury should be potentially liable to pay damages to that person for that injury. These views are plausible, even if open to challenge: there seems to be strong reason not to make A liable to B if A is not responsible for B's injury, and individual responsibility seems to require causation.[3]

At any rate, the common law of tort reflects this position: the general rule in tort law is that a claimant must prove on the balance of probability that a particular defendant's wrongful conduct was a factual cause of its injury in order to be entitled to compensatory damages in respect of that injury. As a matter of substantive law, this requires that but for the defendant's wrongful conduct the injury would not have occurred or that the wrongful conduct made a material contribution to the injury.[4] As a matter of proof, the claimant must show this on the balance of probability. This itself comprises two rules: the burden of proof, which dictates that the claimant loses in the event that proof on the balance of probabilities is impossible and the standard of proof, which dictates that a fact be found when it is more probable than not.

*Fellow and Associate Professor of Law, Wadham College, Oxford University. I would like to thank Jane Stapleton, Nick McBride, Ariel Porat, Sarah Green and the anonymous reviewers for very helpful comments on an earlier draft. The usual disclaimer applies.

1 S. Kagan, 'Causation, Liability, and Internalism' (1986) 15 *Philosophy and Public Affairs* 41, 41.
2 Not a conclusive reason and not the only limit, of course.
3 For the view that individual outcome responsibility requires causation, see J. Coleman, 'Doing Away with Tort Law' (2008) 41 *Loyola of Los Angeles Law Review* 1149, 1161–1165. For an argument that causation is not necessary for individual moral responsibility for an outcome, see C. Sartorio, 'Causation and Responsibility' (2007) 2 *Philosophy Compass* 749, 758–759.
4 On why 'material contribution' does not always collapse into the but-for test, see S. Steel, 'Causation and Scope of Responsibility' in K. Oliphant (ed*), Law of Tort* (London: Lexis Nexis, 2015) at [14.12].

Published by John Wiley & Sons Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA

Although the general burden of proof rule has been repeatedly affirmed in English law, exceptions to it have been recognised over the past fifty years.[5] These exceptions allow proof that the defendant's wrongful conduct materially increased the risk of the claimant's injury to generate liability in certain contexts, most notably in cases of mesothelioma caused by exposure to asbestos dust.[6] In these contexts, the claimant succeeds despite it being impossible to prove on the balance of probabilities that the defendant's wrongful conduct was a but-for cause of the claimant's injury or that it made a material contribution to the claimant's injury. Such exceptions raise at least three concerns.[7] First, a moral concern: what justifies departing from the principle of individual responsibility reflected in the general rule, so that even proof on a balance of probabilities of that responsibility is no longer required? Second, a concern of rational limitations. Once proof of causation on the balance of probabilities is no longer required in one context, the concern is that an exception cannot be consistently limited to that context and so, if limits are to be found, they must be arbitrary.[8] This, in turn, gives rise to a third concern – namely, that the absence of rational principles leads to uncertainty in the law.

The existing English law on exceptions to proof of causation fails to do justice to these concerns.[9] Beyond an appeal to unstructured intuitions of fairness, the moral concern has not been adequately addressed.[10] Nor is it clear how one can rationally explain the limits of the current exceptions. A claimant is relieved of proving causation on the balance of probabilities only if causation is impossible to prove due to uncertainties in scientific knowledge (merely factual impossibility of proof will not do) and only if the possible causes of the claimant's injury operate in a substantially similar way to cause the injury in question.[11] The rationality of these limitations upon the exceptional rules has rightly been doubted.[12] Finally, it is far from clear when causal uncertainty is genuinely due to gaps in 'scientific' knowledge or when possible causes should be said to operate in similar ways.

5 For a recent affirmation of it, see *Sienkiewicz* v *Greif (UK) Ltd* [2011] UKSC 10 (*Sienkiewicz*) at [16] *per* Lord Phillips: 'It is a basic principle of the law of tort that the claimant will only have a cause of action if he can prove, on balance of probabilities, that the defendant's tortious conduct caused the damage in respect of which compensation is claimed'.

6 The central modern authority is *Fairchild* v *Glenhaven Funeral Services Ltd* [2002] UKHL 22. For a detailed overview of the current law, see Steel, 'Causation and Scope of Responsibility' n 4 above at [14.20]-[14.27].

7 For all of these concerns in relation to existing exceptions to proof of causation in English law, see eg, Lord Hoffmann, '*Fairchild* and After' in A. Burrows, D. Johnston and R. Zimmermann (eds), *Judge and Jurist: Essays in Memory of Lord Rodger* (Oxford: OUP, 2013) 69, describing the exception in mesothelioma cases as 'an arbitrary rule masquerading as a principled decision, capable of causing endless difficulties as to the limits of its application'.

8 See J. Morgan, 'Causation, Politics, and Law: The English – and Scottish – Asbestos Saga' in R. Goldberg (ed), *Perspectives on Causation* (Oxford: Hart, 2011) 62–63.

9 For a more detailed vindication of these claims, see S. Steel, 'Causation in English Tort Law: Still Wrong After All These Years' (2012) 31 *University of Queensland Law Journal* 243, 246–259.

10 See, again, Morgan, n 8 above, 61: '[*Fairchild*] does little to identify or justify the making of an exception'.

11 *Sanderson* v *Hull* [2008] EWCA Civ 1211 at [52]; *Fairchild* v *Glenhaven Funeral Services Ltd* n 6 above at [22] *per* Lord Bingham; and at [170] *per* Lord Rodger.

12 See, eg, R. Stevens, *Torts and Rights* (Oxford: OUP, 2007) 51; A. Beever, *Rediscovering the Law of Negligence* (Oxford: Hart, 2007) 484. See also, below, 746.

The aim of this article is to show that it is possible to recognise a set of exceptions to the general rule which meets these concerns and avoids the arbitrary lines currently drawn by English law. Three exceptional rules will be defended. Each of them, I claim, is justified as a means of enforcing what the defendant's secondary moral duty requires in the circumstances of the case. The first part of the article explains this idea – the idea of what I will call a 'secondary-duty-based exception' and defends the three exceptional rules. The upshot will be that exceptions can be justified, on this rationale, only where the defendant has wrongfully caused injury to an indeterminate claimant, where the claimant has been the victim of a wrongfully caused injury at the hands of one or more defendants who may have wrongfully caused that injury, and in cases which bear both of these features. In short, the article defends the possibility of exceptions being made in what are sometimes termed 'claimant indeterminacy' and 'defendant indeterminacy' situations, as well as in situations involving a combination of these features.[13] The second part of the article demonstrates how acceptance of the secondary-duty-based rationale for these exceptions would improve the current law and meet the three concerns identified above. In the final part, I offer some arguments against alternative justifications of further exceptional rules, justifications which would lead to exceptions of broader (and more uncertain) scope than those defended here.

Two preliminaries are in order. First, this article defends exceptions to the burden of proof – the rule that the claimant loses in the event that it is impossible to prove on the balance of probability that the defendant's wrongful conduct was a cause of its injury. It is not therefore concerned with the issue of whether, as a matter of substantive law, the law of tort should exceptionally recognise non-causal forms of liability: liability which could be imposed where it is proven that the defendant *did not* cause the claimant's injury. Second, the article is only concerned with exceptions to proof of a causal connection between a defendant's wrongful conduct and the initial injury or loss which grounds the claim for compensation from the defendant. Thus, it is not concerned with issues of quantification nor with the proof rules governing the issue of whether, though the defendant's wrongful conduct was a cause of the claimant's injury, that injury would nonetheless have occurred had the defendant behaved non-wrongfully.[14]

13 See, eg, R. Delgado, 'Beyond *Sindell*: Relaxation of Cause-in-Fact for Indeterminate Plaintiffs' (1982) 70 *California Law Review* 881, 882–883.

14 Stapleton describes this as the issue of whether the claimant has suffered 'damage'. See J. Stapleton, 'Unnecessary Causes' (2013) 129 LQR 39, 54–62. The defendant may also, however, plead that 'it would have happened anyway had I behaved non-wrongfully', where this denies that a wrongful *aspect* of the defendant's wrongful conduct was causative of the claimant's injury. For instance, in German law, the defendant bears the burden of proof of showing that the *wrongful aspect* of unlawful conduct which led to injury, eg, driving at 80 rather than 50, was *not* causative of the claimant's injury. In effect, the defendant's conduct is treated as a whole rather than attention being paid to a single aspect of it. See H. Oetker, *Münchener Kommentar zum BGB*, 4th ed, 2012, §249 BGB, [224]. For an illuminating analysis of *McGhee* v *National Coal Board* [1973] 1 WLR 1 as a case where the defender's conduct should be considered as a whole and, as such, seen as wrongful, and thus established to have been a cause of the pursuer's injury, see E. Weinrib, 'Causal Uncertainty' 18–23, advance access version at http://ojls.oxfordjournals.org/content/early/2015/07/02/ojls.gqv020.full (last accessed 13 July 2015). This kind of analysis can neatly avoid many causal indeterminacy situations without the need for an exceptional rule. For an

SECONDARY-DUTY-BASED EXCEPTIONS TO PROOF OF INDIVIDUAL CAUSATION

Tort law imposes legal duties on people to act or refrain from acting in certain ways for the benefit of other people. Breach of these 'primary' duties gives rise to 'secondary' legal relations between the duty-bearer and the correlative right-holder.[15] The fact that such secondary legal relations arise can partly be justified by the fact that, as a moral matter, the breach of a primary moral duty gives rise to secondary moral relations between the parties. It will typically give rise to a duty to apologise. It will also typically give rise to a moral duty to take steps to compensate the victim for injury caused. Such moral relations come into existence as a rational reflection of the continued force exercised by the reasons which underpinned the defendant's original primary duty.[16] The duty breached, the defendant must take steps to make the world as if the breach of duty never occurred. That is the best that can be done *ex post* to conform to the reasons for the original duty.

Wrongful causation of injury is generally a necessary trigger for the coming into existence of such (ie, those relations giving rise to duties of compensation) secondary moral relations. So how could there be a secondary-duty-based[17] justification for departing from proof of causation, given that such duties of compensation only arise where the defendant has caused the claimant's injury?

The reason is that there might be secondary duties to participate in schemes whereby one can ensure or substantially increase the probability that one will compensate one's victim, even if one's victim cannot be identified.[18] For instance, suppose that D1 and D2 have each wrongfully caused injury to one of C1 and C2, but it cannot be determined which defendant injured which claimant, though each may have injured either C1 or C2. Both of these defendants know that they have a secondary duty owed to *one of C1 and C2*. That duty could itself give rise to an obligation on the part of the defendants to enter into an agreement whereby each agrees to compensate the other's victim. In this way, by authorising another person to pay damages on one's behalf, one discharges one's secondary obligation to one's victim, even though one cannot identify one's victim. As Tadros has written: 'If [a person] cannot fulfil her duty herself she has a duty to get someone else to help her. That is familiar from other kinds of duties. If I promise you to *v* and I cannot *v* myself I typically have a duty

arguable (implicit) application of such a rule, see *Phethean-Coles* v *Hubble* [2012] EWCA Civ 349. cf *The Empire Jamaica* [1955] 1 All ER 452.

15 So much is uncontroversial (even if it is controversial whether *all* liability in tort has this structure). There is controversy over whether the secondary legal relation is a legal duty or a legal liability. I remain non-committal here. On this, see S. Smith, 'Duties, Liabilities, and Damages' (2012) 125 *Harvard Law Review* 1727.

16 See J. Gardner, 'What is Tort Law for? Part 1: The Place of Corrective Justice' (2011) 30 *Law and Philosophy* 1, 28–37.

17 Here I will refer to 'duties' rather than 'liabilities' purely for convenience and subject to n 15 above.

18 For a similar suggestion, see A. Porat and A. Stein, *Tort Liability Under Uncertainty* (Oxford: OUP, 2001) 133, 135–136.

to get someone else to v'.[19] My suggestion is that this type of reasoning can justify the following three exceptional rules.

The first rule is this:

> *Causative defendant indeterminacy rule* (CDIR). If D1, D2 . . . D*n*, each acting independently of the others, have each wrongfully injured one of C1, C2 . . . C*n*, but it cannot be determined in any case on the balance of probability which D has wrongfully injured which C, though it is true of each D that it may have caused each C's injury, then, each D is liable to contribute to a fund in the amount of the injury that it has, on the balance of probability, caused.

Here is a simple case to which this rule would apply:

> *Two cars*. D1 and D2 are hitwomen independently employed to destroy C1's car. Each places a small explosive device near C1's car. The devices explode, but only one device destroys C1's car, whilst the other device destroys C2's car. It cannot be determined which D's device destroyed which car.

Under the general rule, no claimant would be able to establish the liability of either defendant. Each defendant could say: 'you have not proven that I caused your injury on the balance of probability – there is an equal probability that it was the other defendant'. Intuitively, this seems highly unjust.

Here is one explanation of this intuition. It is true of each D that it has either caused a C's injury *or* prevented the C being able to enforce its secondary rights in respect of that injury.[20] It may have prevented this enforcement because its wrongful conduct makes it impossible to attribute causality to a particular defendant. This prevention deprives C of the compensatory damages it would have obtained from the defendant who actually caused the damage to the car. Hence, *each* D should be liable to *each* C for at least the amount of the value of the car.

The fact that each D has either caused injury to C or prevented its being able to enforce its rights in relation to that injury certainly appears to have moral significance, since it can then be said that each D has caused C injury. But to see the limits of this argument, consider the following case:

> *Two causative hunters, different bullets*. D1, D2 and C1, C2 are on a hunting trip. Conscious that an accident might occur, and anxious that responsibility for that accident be capable of determination, D1 and D2 use bullets of different calibres. D1 and D2 negligently fire one shot each and C1 and C2 are each injured by a single bullet. The bullets are retrieved but, maliciously, a hospital employee, X, disposes of the bullets and it cannot now be determined which D injured which C.

Given that X deliberately disposed of the bullets, it is difficult to say that each D may have caused it to be the case that a claim cannot be enforced against the other defendant. Unlike in *Two cars*, there has been a break in the chain of causation

19 V. Tadros, *The Ends of Harm* (Oxford: OUP, 2012) 193.

20 See further, below, 738–741.

between each D's action and C's inability to enforce its claim. Yet it still intuitively seems that D1 and D2, who have certainly each injured one of C1 and C2, could have little complaint against being required to pay compensation. Hence, the argument that each D has caused injury or prevented C being able to enforce a claim in respect of that injury cannot be the full explanation of this intuition.

Here, then, is another explanation. Each defendant is under a secondary moral duty to pay compensation to one of the two victims in *Two cars* and *Two causative hunters, different bullets*. Given that each defendant cannot identify the individual to whom it owes that duty, what should it do? Surely, the answer is not nothing. The mere fact that the victim cannot be identified does not release the defendant from its obligation. A reasonable course of action is as follows. Each defendant can enter into an agreement whereby each authorises the other to pay damages on its behalf. For example, the agreement in *Two cars* might run as follows: 'I agree to pay damages to C1 if you pay damages to C2. In the event that I am not in fact under a duty to pay damages to C1, because you injured C1, I authorise you to act on my behalf in paying damages to C2, if you authorise me likewise to act on your behalf'. In this way, each defendant can be assured of actually compensating its own victim. In the event that the defendants do not enter into such an agreement, they are in breach of what their secondary duty to pay compensation to the victim requires.[21]

This suggests that each C should be able to sue each D individually in *Two cars*, by making the following argument. Each C could argue: you ought to pay me compensation because you either wrongfully injured me or wrongfully failed to enter into an agreement whereby you paid me compensation.

If this argument were accepted, however, it could give rise to the possibility that each claimant could recover the entirety of its loss from one defendant. This would have the undesirable consequence that an individual defendant could be held liable for an amount of loss which it could not possibly have caused.[22] For instance, if each car was worth £1,000, then the maximum amount of loss each defendant has caused is £1,000. It would, then, be problematic to hold one defendant liable for £2,000.

The following set-up would solve this problem. The court should require each defendant to pay the damages it owes into a fund held by the court. The justification for this is that it is a means of enforcing the defendants' secondary duties to pay compensation to their victims. By requiring each D to pay the amount of compensation owed to its victim into a fund, the court provides a means by which each D can satisfy its obligation to its victim. In effect, the fund serves the same pooling function as the authorisation agreement which should have occurred between the Ds.

21 This justification is very different to that proposed by those who consider causation to be morally arbitrary and argue that all those who culpably risk harm should be required to pay into a 'fault pool'. Causation is central to the justification in the text. On the idea of a 'fault pool', see J. Coleman, 'On the Moral Argument for the Fault System' (1974) *Journal of Philosophy* 473, 483–488.

22 It is permissible to risk defendants paying too much in damages (if it were not, certainty should be the standard of proof in civil litigation). What is much harder to justify is knowingly requiring the defendant to pay for more loss than it has caused.

The application of CDIR raises some questions. What if, in *Two cars*, the cars have different values? Under CDIR, each defendant could only be required to pay the value of the less valuable car. This is because it is known only that each defendant caused that amount of loss. A defendant could not be under a secondary duty to pay into a fund a greater amount than it has been proven to have caused. However, it may be possible to rely upon a further exceptional rule in relation to the amount of loss which each defendant's wrongful conduct may have caused. For instance, if one car is worth £500 and the other is worth £750, then each defendant is liable under CDIR for £500. Each defendant may be liable to the owner of the car worth £750 for a further £250 if the law permits defendants to be held liable for a loss which their wrongful conduct *may* have caused, even where they cannot be proven to have caused a loss of that amount.[23]

What if, in *Two cars*, one defendant is insolvent or untraceable? In these circumstances, the solvent defendant still knows that it has a secondary duty to pay compensation to one of C1 and C2, but the effect of requiring the defendant to pay damages for the loss caused into a fund to be dispersed to C1 and C2 will be to require the defendant partially to compensate a claimant whom it has not harmed. If the fund is divided equally between C1 and C2, then 50 per cent of the defendant's money will be paid to someone whom it has not harmed. Can the solvent defendant legitimately complain about this? The answer is negative if each claimant instructs the defendant to pay the amount into the fund. This is because then the defendant knows that the victim of its wrongful conduct wants the money to be dispersed in this way. There is no material difference between this case and the claimant simply instructing the defendant to pay the money into a particular bank account. It is no concern of the defendant's whether his victim decides to share its damages with another victim.

Is CDIR an unimportant rule designed to handle fanciful scenarios such as those already presented? No, consider:

> *Drugs*. D1 and D2 have manufactured several thousand defective medicinal drugs. This has led to hundreds of injuries to hundreds of claimants. But, due to the long latency period between ingestion of the drug and the appearance of the injury, it is impossible to determine (due to lost records, faded memories, etc) which D caused which C's injury. It is known however that each D had a 50 per cent share of the national market at the time each C ingested the drug.

Such cases have arisen in the United States and, more recently, in France.[24] They are not uncommon or fanciful. Under CDIR, each D should pay for 50 per cent of total loss caused by the defective drugs, on the assumption that the national market share is a reasonably accurate reflection of the amount of harm the D has caused. Notice that the justification for this is not an idea of aggregate or collective corrective justice.[25] The justification is rather that the best way the

23 Such a rule is defended below: 737–742.

24 *Sindell* v *Abbott Laboratories* (1980) 26 Cal. 3d. 588. For the French cases, see S. Steel, '*Sienkiewicz v Greif (UK) Ltd* and Exceptional Doctrines of Natural Causation' (2011) 2 *Journal of European Tort Law* 294, 305–308.

25 Compare R. Wright, 'Liability for Possible Wrongs: Causation, Statistical Probability, and the Burden of Proof' (2008) 41 *Loyola of Los Angeles Law Review* 1295, 1326–1328.

defendant can comply with its secondary duty towards its victim is by paying into the fund. It is individualised or relational corrective justice.

Nor does the justification depend upon all of the victims or other tortfeasors being joined to the proceedings. Each defendant has a secondary duty to authorise other defendants to act on its behalf so long as this raises the probability that the defendant will comply with its secondary duty to a standard higher than the balance of probabilities. Thus if there were three defendants, each of whom has caused wrongful injury to one of three claimants, but only two claimants are joined to the proceedings, each defendant is still required to pay into the fund. This is because there remains a 2/3 probability that the defendant will comply with its secondary duty by paying into the fund. Similarly, if only two of the defendants were joined to the proceedings, each defendant can still be required to pay into the fund. This is because, by entering into an authorisation agreement with the other defendant, each defendant would have increased the probability of complying with its secondary duty from (at least) 1/3 to 2/3.

Now I will defend a second rule:

> *Indeterminate claimant rule*. If D has wrongfully caused injury to one of a group of people, C1-C*n*, each of whom may have been caused injury by D, but it cannot be determined which C has been injured by D on the balance of probability, D should be liable to pay the amount of loss D has caused into a fund to be dispersed between C1-C*n*.

Here is a case to which this rule would apply:

> *Chancy antidote*. A venomous snake bites three people. An ambulance is called. Due to negligence, the ambulance does not turn up and all three suffer serious injuries due to the lack of antidote. Had the ambulance turned up, it would not have had enough antidote to treat all three people, but only one person. The decision as to who would have received the antidote would have been taken by lot.

Here one person would not have suffered wrongful injury had the ambulance turned up, but it is not known which.[26] As things stand, each person can show there is a 1/3 probability that it was wrongfully caused injury by the ambulance's negligence. Consequently none would recover damages under the general rule.[27]

Under the *indeterminate claimant rule*, each would recover 1/3 of their loss. This can be justified in two ways. First, if each claimant agrees that the defendant should pay into the fund, then, again, the defendant knows that it is complying with its actual victim's demands. It is just as if its victim asked the defendant to pay the money into a particular bank account. In responding to its victim's

26 Let us assume that it was not negligent to have turned up with insufficient antidote.

27 Nor would they, as the law currently stands, recover for the lost chance of avoiding their physical injuries. This follows from *Hotson* v *East Berkshire Health Authority* [1987] AC 750. For reasons explained in S. Steel, *Proof of Causation in Tort Law* (Cambridge: CUP, 2015, forthcoming) ch 6, it is preferable not to develop the scope of substantial compensatory damages for lost chances any further.

demands, the defendant is simply conforming to its moral duty to pay compensation to that victim. Second, more controversially, someone who wrongfully injures a person comes under a non-relational (a duty not owed to anyone in particular) moral duty to contribute to the relief of suffering. Suppose that D wrongfully kills C, who has no dependants. D must still try to do the next best thing to conforming to its original duty not to kill C. Any such action cannot benefit C. Consequently it is difficult to describe any secondary duty as a duty owed to *C*. But surely D has some kind of secondary moral duty in the face of his wrongdoing. The best way for D to respond to his wrongdoing is to contribute to the relief of suffering. If the law chose to hold each D liable in *Chancy antidote*, this could also be justified on the ground that, in relation to the two Cs who are not the victims of D's wrongdoing, D is fulfilling its duty to assist in the relief of suffering.

Consider now a third rule:

> *Defendant indeterminacy rule*. If C has been the victim of a tortious wrong at the hands of one or more of a group of defendants, D1-D*n*, but it cannot be determined on the balance of probability which defendant(s) amongst that group was a cause of C's injury, but each defendant may have been a cause of C's injury, then each D should be liable to C in full or proportionate to the probability that D was a factual cause of C's injury.

The classic situation to which this rule applies and to which it has been applied is:[28]

> *Two hunters*. Independently of each other, D1 and D2 negligently fire in the direction of C whilst on a hunting trip in the woods. C is injured by one of D1 and D2, but it is impossible to determine on the balance of probability which D caused C's injury.

French, German, Canadian, and the law of virtually all US States would hold each D liable in full in this situation.[29] English law would probably hold each D liable in proportion to the 50 per cent probability that it caused C's injury.[30]

28 The case is based upon *Summers* v *Tice* (1948) 33 Cal. 2d. 80 and *Cook* v *Lewis* [1951] SCR 830.

29 See Steel, n 24 above, 302–308. For Canadian law: see *Cook* v *Lewis, ibid*, approved in *Clements* v *Clements* 2012 SCC 32.

30 See the remarks of Lord Phillips in *Sienkiewicz* v *Greif UK Ltd* [2011] UKSC 10, [105], read in the light of *Barker* v *Corus (UK) Ltd* [2006] 2 AC 572. That proportional liability under *Barker* remains the position at common law (that liability then being rendered *in solidum* by virtue of Section 3 of the Compensation Act 2006 in mesothelioma cases) is confirmed by the recent Supreme Court decision in *Zurich Insurance Plc UK Branch* v *International Energy Group Ltd* [2015] UKSC 33 (see especially at [29]-[33] *per* Lord Mance). Furthermore, proportional liability under *Barker* has held to apply beyond mesothelioma cases: see the decision in *Heneghan* v *Manchester Dry Docks Ltd* [2014] EWHC 4190 (applying *Barker* to a case involving adenocarcinoma of the lung). For discussion of the decision in *Heneghan*, see S. Steel, 'On When *Fairchild* Applies' (2015) 131 LQR 363.

How can this rule be justified? Let us set aside an unsuccessful justification. It has been argued that C can prove that it would not have suffered its injury but-for the negligent actions of D1 and D2, considered as a group.[31] Hence but-for causation is in some sense satisfied. There are two failings here. On the one hand, some justification is needed for grouping together two separate legal entities that are not responsible for each other's actions.[32] This argument itself provides no such justification. On the other hand, the fact that, but-for the wrongful conduct of a number of persons, where the defendant is one of that number, the claimant would not have suffered an injury, itself provides only very weak reason to impose liability on a particular defendant. Suppose D1 wrongfully fires and strikes C and that had D1 not fired and struck C, D2's bullet would have (the effect of D1's bullet is to cause C to collapse before D2's bullet arrives). But-for the negligent actions of D1 and D2, C would not have suffered injury. But this provides little or no reason for imposing liability upon D2 – though it is true that but-for the wrongful conduct of D1 and D2, C's injury would not have occurred.[33]

Beever has argued that, in *Two hunters*, each defendant violates the claimant's right to bodily integrity.[34] The argument seems to be as follows: each defendant violates C's right to bodily integrity either by actually causing C physical injury *or* by preventing C being able to enforce its right to damages. One defendant prevents C being able to claim damages since, if that defendant had not fired, it would have been possible for C to sue the defendant who caused the injury, since that defendant would then be the only one who fired. Beever claims that preventing C being able to enforce C's right to bodily integrity is itself a violation of that right, where D acts in such a way that would generate liability for causing C physical injury if it were known that D had caused the physical injury (in other words: where D acts negligently with respect to C's bodily integrity). Let's call this the *prevention argument*.

Before we can assess this, some clarification is in order. It is not clear what it means to say that C has a 'right to bodily integrity'. In English law, C has a primary legal right that each D not by its negligent action cause reasonably foreseeable physical injury to C's body. The infringement of that right by a defendant gives rise to a secondary duty or liability to pay compensatory damages as between the defendant and victim. So C has two rights 'to' bodily integrity: a primary right in relation to D's conduct and a secondary right arising out of the violation of that primary right.

Once unpacked in this way, it becomes glib to say that each D has invaded C's bodily integrity. It is only true of each D that it has either infringed C's primary right not to be negligently physically injured or negligently prevented C from

31 *Clements* v *Clements* 2012 SCC 32 at [39]-[41].

32 For a clear explanation (and extended justification) of this point in a related context, see J. Stapleton, 'An Extended But-for Test for the Law of Obligations' 21–24, advance access version at http://ojls.oxfordjournals.org/content/early/2015/03/06/ojls.gqv005.full.pdf (last accessed 22 May 2015).

33 Compare S. Green, *Causation in Negligence* (Oxford: Hart, 2014) 72–73.

34 A. Beever, n 12 above, 459–464. Versions of this argument have been made since the late nineteenth century in French and German law: see Steel, n 27 above, ch 4.

enforcing its secondary legal relations against the other defendant. Preventing the repair of a right's infringement is not the same as infringing the right.[35]

How, then, does Beever bridge the analytical gap between primary and secondary rights, between preventing the repair of a right and actually infringing the right? This matters because there is, on the face of it, a significant difference between a defendant's being negligent in relation to the claimant's bodily integrity, and the defendant's being negligent in relation to the claimant's right to compensation in relation to its bodily integrity. An obvious problem with the argument is that whilst each defendant has behaved negligently with respect to the claimant's actual physical integrity, neither has behaved negligently with respect to the interference with its secondary right to physical integrity. In other words, the interference with the secondary right is a remote consequence of the defendant's negligence in relation to the claimant's bodily integrity.

Beever's argument seems to rely upon two ideas. One is that compensatory damages paid by a defendant in respect of a physical injury to the claimant in normative terms repair the physical injury. Hence, when the defendant prevents the claimant obtaining compensatory damages it is the same as actually physically injuring the claimant. The other is that the claimant's (secondary) right to compensation is one and the same right as its primary right to bodily integrity;[36] as a matter of normative reality C's secondary right is just a normative continuation of C's primary right.[37] Hence, when D prevents C enforcing its secondary right it is *ipso facto* infringing C's primary right so long as D otherwise satisfies the requirements for infringing C's primary right – namely, negligence in relation to C's bodily integrity.[38]

This ingenious line of reasoning cannot be accepted in this form. On the first argument, it is simply fictitious to equate prevention of a claim for damages in respect of a physical injury with physical injury. We need a reason for this equation. On the second argument, C's legal right to compensatory damages in respect of negligence is different from C's right that D not negligently cause reasonably foreseeable injury to C's body. The duties correlative to these rights have entirely different contents.[39] Moreover, secondary rights can sometimes be assigned in situations where the corresponding primary right cannot. This tends to suggest that these rights are not identical.

35 Not paying for the vase to be fixed is not breaking the vase.

36 Beever, n 12 above, 461: 'The argument is that [one of the hunters] violated [the other's] primary right to bodily integrity, though he did not . . . harm [his] body. The argument does not rely on any new right, but on the one that always existed: the right to bodily integrity'. This argument is more prominent in Weinrib's account of the *Two hunters* case: see E. Weinrib, 'Causal uncertainty' n 14 above, 10–14.

37 See also, for an extended statement of this general view, E. Weinrib, *Corrective Justice* (Oxford: OUP, 2012) ch 3.

38 Beever, n 12 above, 461–462. At least, this seems to be the implicit premise of the view which equates the non-repair of a right with its infringement.

39 This is perfectly consistent with the view that the reasons underpinning the right continue to exert force after the right's violation. But the fact that the original reasons hang around demanding conformity does not mean that any obligation generated by those reasons is the same obligation as previously existed.

Nonetheless, it seems that this line of reasoning is ultimately defensible.[40] The crucial premise that is in need of defence is that it is permissible to equate negligence in relation to the claimant's bodily integrity with negligence in relation to the claimant's secondary right to compensation in respect of thereof. The Beever argument tries to establish this premise by presupposing an analytical identity between primary and secondary rights. But the argument might still be successful even if secondary rights are only highly similar to primary rights. This much is true: '. . . an award of damages for clinical negligence is, in a sense, the equivalent of proper clinical treatment: it is the nearest the law can get to putting the patient in the position that she would have been if the doctor had not been negligent'.[41]

It might still be objected that behaving negligently in relation to someone's bodily integrity (a physical thing) is significantly different to behaving negligently in relation to a the claimant's secondary right (a purely normative entity), with the result that negligence in relation to the former cannot be taken as negligence in relation to the latter.

A response to this begins with an appeal to the moral function of remoteness requirements. Why do we insist that the damage caused by the defendant's wrongful conduct be of a kind foreseeably risked by that conduct? One plausible answer is that the injury suffered must be among the reasons why the conduct in question was wrongful in order for it to generate a secondary duty to pay compensation. In other words, the requirement of remoteness falls out of the thesis that the reason(s) for secondary obligations of compensation arise out of the reasons for the primary obligation.[42] If the origin of secondary duties of compensation in respect of a harm lies in the fact that this harm was one of the reasons why the conduct was wrongful, then this harm must indeed form one of the reasons why the conduct was wrongful in the first place. But if the harm is not foreseeably risked by a certain action, then it does not form part of the reasons which made that action wrongful. The reason carelessly risking physical injury is wrongful is that it may cause physical injury. The secondary right to compensation, in turn, arises as a reflection of that reason. Consequently, there is a sense in which acting wrongfully in relation to a person's physical integrity is acting contrary to the reasons which give rise to a secondary right in relation to that physical integrity.[43]

Even if this response is not accepted, it seems that in many of the cases which will fall within the *defendant indeterminacy rule*, such as the *Two hunters* case, the interference with the secondary right is itself, ie described in those terms, a reasonably foreseeable consequence of the defendant's wrongful conduct. It does

40 The revision here to the account offered by Beever (and to the account offered by Weinrib, n 36 above) is minor. My account is much indebted to their ingenious efforts.

41 *Wright* v *Cambridge Medical Group* [2011] EWCA Civ 669 at [60] *per* Lord Neuberger MR.

42 See n 16 above.

43 Yet I do not know what to say about this case, where it seems that the loss must be too remote: *Missed tort case*. D1, a bus driver, is driving C to court for C's tort case against D2 in respect of a physical injury D2 has negligently caused C in a car accident. D1 drives negligently, risking C an injury of exactly the same kind as D2 has caused C, and is involved in a collision. Though physically uninjured, C is unable to make it to court on time and the case against D2 is dismissed.

not seem unreasonable to say that people who negligently fire bullets in close temporal and spatial proximity ought to be aware that this might lead to difficulty in determining who caused an injury to nearby individuals.[44]

In sum, then, it seems that the *prevention argument* has force and can justify holding each D liable in full in situations like *Two hunters*. Unfortunately, however, the *prevention argument* has limits. It is doubtful whether it could justify liability in all cases which might fall within the *defendant indeterminacy rule* as formulated. First, it is doubtful whether it could justify liability in:[45]

> *Two hunters, different bullets*: D1, D2 and C are on a hunting trip. Conscious that an accident might occur, and anxious that responsibility for that accident be capable of determination, D1 and D2 use bullets of different calibres. D1 and D2 behave negligently and C is injured by one bullet. The bullet is retrieved but, maliciously, a hospital employee, X, disposes of the bullet and it cannot now be determined which D injured C.

In this case, the causal link between the defendant who negligently prevented the enforcement of C's secondary right against the appropriate defendant has been broken by the malicious intervention of X. Consequently, it is not true of each D that it either caused the physical injury or prevented the enforcement of the claimant's secondary legal right. Though one D is a but-for cause of the prevention of the enforcement of the claimant's secondary legal right, it is not typically a *legal* cause, given the fact of the malicious intervention, which would break the chain of causation between that defendant and the interference with the secondary right.

Second, if one D is insolvent in *Two hunters*, the *prevention argument* faces a problem. It is then not true of each defendant that it is at least liable for the amount of the claimant's damages since, if the insolvent defendant caused the

44 Any objection that the loss is purely economic is already met by the high similarity between the primary and secondary rights in respect of a person's physical integrity. One might also question whether the general rule concerning pure economic loss is not justified mainly by considerations relating to the generation of indeterminate liability. In the *Two hunters* situation, this is not a live consideration. Each D is only being held liable for the amount of damages it would have been held liable for had it been known that its negligence had caused the physical injury.

45 It might seem that the prevention argument has another limit. Does it only apply to cases involving two wrongful actions? If, however, there had been three (or more) hunters who negligently fired in the direction of the claimant, the outcome of the analysis would not change. Even if it is then not true that one of the hunters is a but-for cause of the claimant's inability to enforce its secondary right, it will be a NESS (a **n**ecessary **e**lement for the **s**ufficiency of a **s**et) cause of this (on NESS causation, see generally: R. Wright, 'Causation in Tort Law' (1985) 73 *California Law Review* 1735). In other words, the inability to enforce will be causally overdetermined by the actions of two of the hunters. It might then be objected that neither of those hunters has rendered the claimant any worse off, since it would have been unable to enforce its claim in any event due to the action of the third hunter. But this argument, which involves direct reliance upon a third party's wrong to avoid a liability which would otherwise arise, is blocked by *Baker* v *Willoughby* [1970] AC 467. Alternatively, one could argue that each hunter has either: (1) wrongfully caused the physical injury, (2) prevented the enforcement of the secondary right in respect of that physical injury or (3) prevented the enforcement of the claim in respect of the prevention of the enforcement of the secondary right in respect of the physical injury. Each hunter is liable in full regardless of which of (1)–(3) is in fact the case.

physical injury, then preventing a claim against that defendant is to prevent a valueless claim. To avoid this, one would need Beever's conceptual fiction that preventing the repair of a right is identical to infringing the right.

Two further justifications of the *defendant indeterminacy rule* avoid these problems, but at the cost of depriving the rule of a familiar secondary duty-based justification. The first is that there is something normatively problematic about the causative defendant relying upon the wrongful conduct of another person to deprive the claimant of a right to compensation in respect of an injury caused by wrongful conduct.[46] On the one hand, this allows the defendant to benefit from another person's wrongful conduct. On the other hand, this deprives the claimant, who has suffered injury due to wrongful conduct, of a right to compensation. This normative concern is found elsewhere in the law of tort. For instance, a defendant who has wrongfully injured a person is prevented from pointing to the fact that the wrongful conduct of a third party would have caused the same consequential loss in any event in order to reduce their liability.[47]

The second justification of the *Defendant indeterminacy rule* runs as follows. If no liability is imposed in *Two hunters* there is a 100 per cent risk of injustice falling upon one person – the claimant. If liability is imposed then each defendant is exposed to a 50 per cent risk of injustice. Although there is, across persons, a 100 per cent risk of injustice regardless of whether liability is imposed or not, if no liability is imposed, one individual is exposed to the full 100 per cent risk, whilst if liability is imposed, the risk of injustice is shared between two individuals. This seems to be a fairer distribution of the risk of injustice.[48]

An objection: paradox[49]

A premise of the prevention argument is that one defendant has prevented the claimant from being able to recover damages in respect of the infringement of its primary right by the other defendant. But this premise seems to be contradicted by the fact that, if the prevented claim theory is successful, the claimant can after all sue the defendant who caused the physical injury. Thus, it turns out that no claim has been prevented after all. But, if the claimant has not been deprived of a claim, then the basis of the defendants' liability collapses. If so, then the claimant cannot sue either defendant, and has been deprived of a claim. Then it follows that the claimant can sue both defendants. But if it can sue both, then it has not been deprived of claim against the defendant who caused the physical injury. It seems that the law is propelled into an infinite regress here.

46 For a more detailed exposition of this justification, see S. Steel, 'Causation in English Tort Law: Still Wrong after All These Years' (2012) 31 *University of Queensland Law Journal* 243, 260–263; Steel, n 27 above, ch 4.

47 *Baker* v *Willoughby* n 45 above. See also *Haxton* v *Philips Electronics UK Ltd* [2014] EWCA Civ 4, where the defendant was prevented from relying on the wrong it committed against a third party to reduce its damages to the claimant.

48 Compare G. Williams, (1953) 31 Can BR 315, 317.

49 Thanks to Donal Nolan and Rob Stevens for making clear to me that this objection applies in relation to the *Two hunters* case. For a statement of the objection, see Stevens, n 12 above, 150. Compare N. J. McBride and S. Steel, 'Suing for the Loss of the Right to Sue: Why *Wright* is Wrong' (2012) 28 PN 27, 37–38.

The paradox objection can be defused by more precisely specifying the claim that the claimant has been prevented from pursuing. If the claim which the claimant is given under the prevented claim argument against the defendant who caused the physical injury is different to the claim that it was deprived of against that defendant by the other defendant, then the paradox does not arise. The paradox only arises because it seems that the result of applying the theory is that the claimant is not deprived of a claim against the defendant who caused the physical injury after all – since it can sue both defendants. And it is true that the claimant is not deprived of 'a' claim against that defendant. But the prevented claim argument does not assert this: it asserts that the claimant was deprived of its ordinary physical injury claim (which would have been available on the balance of probability) against the causative defendant. The claimant *has* lost *this* claim. The claim arising from the prevented claim theory is not the same claim as the one of which the defendant deprived the claimant.

THE EFFECT OF THE PROPOSED EXCEPTIONS ON THE CURRENT LAW

Suppose the law accepted:

(1) the causative defendant indeterminacy rule
(2) the claimant indeterminacy rule *and*
(3) the defendant indeterminacy rule

Would it be necessary, as a matter of consistency, to reject the general rule that the claimant prove causation on the balance of probability in order to be entitled to substantial compensatory damages?[50] No, consider the commonest type of case of causal uncertainty:

> *Simple causal uncertainty*. C cannot prove on the balance of probability that D's wrongful conduct was a cause of C's injury where, if D's wrongful conduct has caused injury, it will have been C's injury, and if C's injury has been caused by wrongful conduct, it will have been D's wrongful conduct.

Here C cannot prove that (a) D has wrongfully injured anyone nor that (b) C has been a victim of a tort on the balance of probability. Therefore (1) would not apply: it requires both (a) and (b). Nor would (2) apply: it requires (a). Nor would (3) apply: it requires (b) to be satisfied. These restrictions are justified on the following grounds. Unless at least (a) or (b) is satisfied, C cannot claim that liability is justified on the basis of enforcing D's secondary duty to pay compensation. Under (a) this duty will arise through the fact that the defendant has wrongfully caused injury to someone, where that person may have been the claimant. Under (b) this duty will arise on the basis that each D has either wrongfully injured C or caused it to be unable to enforce its rights against the other defendant. In short, the exceptions only apply where C has been the victim

50 Compare, eg, J. Morgan, 'Causation, Politics, and Law: The English – and Scottish – Asbestos Saga' in R. Goldberg (ed), *Perspectives on Causation* (Oxford: Hart, 2011) 60.

of a tort on the balance of probability or where D has committed a tort on the balance of probability.

In what ways would the acceptance of (1)–(3) change the current law? Cases such as *Wilsher* v *Essex AHA*, where the claimant could not prove on the balance of probability that its injury had occurred due to a tort, would not be decided differently.[51] In that case, the claimant baby suffered blindness which was alleged to have been caused by the defendant's negligent over-supply of oxygen to the baby's bloodstream. Though the negligence had increased the risk of the blindness, it could not be proven to have been a cause of it on the balance of probability, since several other non-tortious risk factors for the blindness existed, which may also have caused the blindness. The claimant could not then prove that it had been the victim of a tort on the balance of probability or that the defendant had committed a tort on the balance of probability.

If it truly were impossible to prove causation against any particular defendant on the balance of probability on the facts,[52] the leading case, *Fairchild* v *Glenhaven Funeral Services Ltd*[53] (*Fairchild*) would fall within rule (3) and therefore would not be decided differently. In that case, mesothelioma victims or their representatives sued their former employers, who had negligently exposed them to asbestos dust. It was impossible due to the limits of scientific knowledge to attribute any victim's mesothelioma to a particular negligent employer on the balance of probabilities. It was, however, assumed in *Fairchild* that all of the potentially causative factors of the victims' mesothelioma were tortious. Hence, it was taken to be established on the balance of probability that the claimant had been the victim of a tort, only the injurer(s) being unknown.

The subsequent House of Lords decision in *Barker* v *Corus UK Ltd*[54] (*Barker*) is more complicated. In *Barker*, the claimant had suffered three material exposures to asbestos in his working career. Two of them were breaches of duty, whilst one was due to his self-employment. Green argues that one should assign an equal probability of causation to each exposure on the ground that there is no evidence favouring any particular exposure as causative of the mesothelioma.[55] This, in turn, is justified by the 'principle of indifference' – an epistemic principle which holds that an equal probability should be assigned to each member of a set of possibilities, where there is no relevant evidence favouring one possibility over another.[56]

One problem with this argument is that it is inconsistent with the reasoning in *Fairchild* itself. The possibilities of causation in asbestos mesothelioma cases, properly enumerated, are not simply either exposure 1 or exposure 2 or exposure 3 etc, but also include combinations of these exposures. This is because we have no relevant evidence justifying an assumption that the exposures do not operate in synergy to cause the mesothelioma. So, in *Fairchild*, in one case, the victim had been exposed to asbestos by three defendants in breach of duty. The relevant

51 [1988] AC 1074.
52 On why this potentially surprising proviso is necessary, see the immediately following discussion.
53 [2003] 1 AC 32.
54 n 30 above.
55 Green, n 33 above, 69–70, 131.
56 *ibid*, 69–70.

possibilities are as follows (where D1 means 'D1 caused the mesothelioma' and D1D2 means 'a combination of D1 and D2 caused the mesothelioma' etc): D1, D2, D3, D1D2, D1D3, D2D3, D1D2D3. Now, if we assume, following the principle of indifference, that each of these possibilities is equally likely, then the probability that D1 caused the mesothelioma is 4/7. Therefore, on the balance of probability D1 was a cause of the mesothelioma (the same is true of each defendant). In short, if the principle of indifference is applied to *Barker*, the fundamental assumption of *Fairchild* – that it is impossible to prove causation on the balance of probability – is not met.[57]

The question of whether the law should accept the principle of indifference, and if so, whether it truly applies to asbestos-mesothelioma cases, cannot be resolved here.[58] If the principle is accepted and is applicable to the facts of *Barker*, then there is no need for any exceptional rule. The defendant should be held liable *in solidum*, subject to a discount for the claimant's contributory negligence. This would not obviate the need for an exceptional rule in all asbestos-mesothelioma cases, however. As the number of defendants increases, it will not be possible to argue that each defendant was a cause on the balance of probability.[59] In such cases, it could be possible to establish that the victim's injury was caused by a tort on the balance of probability, by virtue of the principle of indifference, even if no individual defendant can be said to be a cause on the balance of probability. In that event, the case would fall within the defendant indeterminacy rule. But even if the principle of indifference is not accepted, it would still sometimes be possible for claimants in asbestos-mesothelioma cases (and, of course, other cases) to come within rule (3). In any case where any environmental asbestos exposure can effectively be discounted, then the claimant will be able to prove that the mesothelioma was caused on the balance of probability by a tort.

On any view as to what suffices as proof on the balance of probability, *Sienkiewicz* v *Greif UK Ltd* would be decided differently.[60] The claimant in this case could not prove that its injury had probably been the result of a tort, rather than environmental exposure to asbestos dust, as the statistical evidence purporting to show this was invalid. Therefore none of the exceptional rules applies: each requires either that the defendant has committed a tort on the balance of probability or that the claimant has been the victim of a tort on the balance of probability. Even if one were to assume that the environmental exposure was *equally* as probable to have caused the injury as the wrongful exposure by the defendant, this would not constitute proof on the balance of probability that the injury was the result of a tort (but only a 50 per cent probability thereof).

57 See also Beever, n 12 above, 477–482.

58 See further, Steel, n 27 above, ch 2. It is doubtful that the principle can be applied to all cases involving asbestos mesothelioma. It is not true that the risks imposed by each exposure, regardless of length, intensity, or type of asbestos dust, are the same. If this were true, it would not make sense to speak of certain kinds of asbestos (eg, blue asbestos) as 'more dangerous' than other types of asbestos dust.

59 I leave the number of defendants at which this becomes true as a diversion for the reader.

60 n 5 above.

Finally, it might be asked why these modifications to the current law would constitute an improvement. Firstly, the main reason is that all of the exceptions have, at an abstract level, a clear and simple justification: in each case the defendant is being held liable as a means of enforcing its secondary duty to its victim. In other words, all of the exceptions are consistent with a highly orthodox justification of liability. Even if we should not accept that all of tort law is justified as a system for enforcing secondary moral duties arising out of the breach of primary moral obligations, this should not obscure the fact that the breach of an important (sub-set) of (relational) moral obligations provides a strong reason for legal intervention. Thus, the moral concern that no clear justification for exceptional rules exists is met.[61]

Secondly, once we realise that exceptions to the requirement of proof by the claimant that a particular defendant was a cause of the claimant's injury are justified by proof that the defendant is under a secondary duty to participate in a system which benefits the claimant, there is no need to insist upon normatively inexplicable limitations upon the existing exceptions. Thus there is no necessity for proof of causation to be impossible for reasons due to the limits of scientific knowledge under any exception. For instance, the CDIR rule applies in the *Drugs* case described above, where the uncertainty was due to the passage of time and absence of records. All that matters is that proof be impossible on the balance of probability through no fault of the claimant.

Similarly, there is no reason the defendant indeterminacy rule should only apply where there is a single potential causative agent.[62] Suppose that two people intentionally, independently, administer two different (they have entirely different chemical compositions) poisons which each expose another person to a 20 per cent risk of death. The poisons operate in different ways to cause injury, but the causal process of each poison, if it does cause death, only lasts a matter of seconds. It is known that only one poison caused the person's death, but it cannot be determined which due to the short timeframe in which the causal processes are active. Here it is true that one defendant's action wrongfully caused the death and one wrongfully prevented the enforcement of the secondary right in respect of that death. Therefore, the defendant indeterminacy rule should apply even though the potential causes have entirely different chemical properties and operate in different ways. The single agent rule is therefore, at best, a rule of thumb for identifying situations where it cannot be determined on the balance of probability which of a number of potential agents were causative of an injury.[63] This is because, if the potential agents are of the *same kind*, then it will

61 See above n 730.

62 This limitation was affirmed in *Barker* v *Corus* n 30 above at [23]-[24] *per* Lord Hoffmann.

63 For further criticisms, see Steel, n 46 above, 248–249; K. Wellington, 'Beyond Single Causative Agents: the Scope of the *Fairchild* Exception post-*Sienkiewicz*' (2013) 20 *Torts Law Journal* 208. For a defence, see Green, n 33 above, 148–151. Green claims that removing the single agent requirement 'would effectively mean imposing liability for risk creation, regardless of whether or not that risk actually resulted in injury', *ibid*, 150. The discussion in the text demonstrates that this is not so. Moreover, even if one rejects the argument in the text and nonetheless accepts liability in *Fairchild* (as Green does) one is still imposing liability for 'risk creation, regardless of whether or not that risk actually resulted in injury': one defendant's wrongful risk imposition may not actually have resulted in injury.

be inherently difficult to show which of them was in fact causative. But, as the example shows, this need not always be the case.

ARE FURTHER EXCEPTIONAL RULES JUSTIFIED?

Several alternative potential justifications for departing from the orthodox requirement that the claimant prove individual causation on the balance of probability can be identified. They can usefully be grouped into four kinds: culpability-based, responsibility-based, incentives-based, and relative-cost-based. In this section, I offer a brief critique of these kinds of argument. I do not claim to offer conclusive objections here to these arguments, but I do aim to cast the burden of proof on those who would defend them.

Culpability

Where a defendant has behaved with a high level of culpability in relation to the victim of an injury, it may be argued that the law should be less concerned to avoid a false positive against the defendant on the causal element of liability. German law requires defendants who have grossly breached a duty to take care not to cause physical injury to another to prove that that gross breach was not a factual cause of injury to a person who has suffered an injury which was culpably risked by that gross breach.[64] Thus if a doctor behaves with gross negligence, it falls to the doctor to prove that his gross negligence was not a cause of an injury suffered by his patient.[65] The idea that the causal requirements of liability might be modified by reference to the defendant's culpability is not alien to English law. In relation to *legal* causation, the remoteness requirements in the intentional torts are less stringent than the requirements in negligence-based torts.[66] Furthermore, it has been held that the defendant can only avoid liability for loss caused by fraud in the tort of deceit by positively proving that loss resulting from that fraud would have happened in any event: the burden of proof for the 'it would have happened in any event' inquiry is upon the defendant.[67] The same is probably true in false imprisonment.[68]

To test the culpability justification, let us take a situation of high culpability and causal uncertainty:

> *Poisoner*. D, C's business rival, attempts to kill C, a critically ill patient, by poisoning C. C suffers paralysis. It is impossible to determine on the balance of probability

64 BGH NJW 1956, 1835, VersR 1956, 499 (medical liability).

65 The rule, initially created by the courts, has recently been codified in §630h (5) BGB. See generally, Steel, n 27 above, ch 5.

66 'The intention to injure the [claimant] disposes of any question of remoteness', *Quinn* v *Leathem* [1901] AC 495, 537.

67 *Downs* v *Chappell* [1997] 1 All ER 426, 433; *BHP Billiton Petroleum Ltd* v *Dalmine SpA* [2003] EWCA Civ 170 at [36].

68 *Lumba* v *Secretary of State for the Home Department* [2011] UKSC 12 at [359].

whether the poison contributed to this paralysis, which was already risked by C's pre-existing critical illness.

Orthodoxy holds that C should not recover compensatory damages in respect of the paralysis because C cannot prove on the balance of probability that D was a cause of his injury. Does D's high culpability justify altering this position? Here are some preliminary reasons against altering the standard position. First, it might be plausibly be maintained that each of us has a (moral) right that a case be proven against us, whether in crime or tort. It seems odd that the more culpably we behave the less we deserve the protection of this procedural right.[69] Second, it is not clear why we should focus on any particular act of culpability in departing from the orthodox burden of proof. Suppose that D has gone unpunished for some previous culpably committed crime. Can it plausibly be said that in virtue of that fact D merits less protection of the law of evidence in a subsequent piece of civil litigation? We need a reason culpability matters to reducing the level of proof needed to render the defendant liable to pay compensation which does not simply amount to depriving bad people of their right to have a case proven against them.

Here is a candidate argument. Since D has behaved with a high level of culpability, D merits punishment. One way of punishing D is to reverse the burden of proof as against D on the causal issue of liability. This inflicts a risk of harm on D to which D would not otherwise be subject, namely, the risk of a false positive on the causal issue. Why might this be apt? Because D has culpably risked an injury to C. Apt retribution for D's culpable risking of injury to C is culpably risking injury to D.

This argument could surely not apply if D had already been criminally sanctioned in respect of his conduct. But let us put that aside. It still seems in need of explanation why we should inflict a punishment on D *by taking away D's evidential rights*. This is one way to expose D to a risk of injury (if we accept the crude retributivism in the argument). But it could be done in many other ways, for instance, by requiring D to pay into a fund for accident victims, dependent upon the outcome of a coin toss. This reveals another potential problem with the argument. Given that D's culpability only seems to impact upon his *desert*, it is not clear why D's desert should benefit C in particular. There seems to be no reason to privilege C in evidential terms over providing some benefit to someone else.[70]

Here is another argument for allowing culpability to reduce the level of proof required for causation or to reverse the burden of proof. Suppose that you have to choose between poisoner being held liable for the paralysis which he did not cause and poisoner not being held liable when he did cause the paralysis. Both involve injustice. Which is the greater injustice? Some people will say that poisoner not being held liable in damages when he did cause the paralysis is more

69 For the seeds of a different view, see A. Porat and E. Posner, 'Aggregation and Law' (2012) 122 *Yale Law Journal* 1, 5–6.

70 This, in substance, is a version of Weinrib's influential 'bipolarity' or 'bilateralism' argument against compensation-deterrence theories of tort law. See E. Weinrib, *The Idea of Private Law* (Oxford: OUP, 2nd ed, 2012) x–xiii, 46–55.

unjust than poisoner being held liable when he did not cause the paralysis. The view is that a false negative is more unjust than a false positive. Perhaps the thought is this. A false positive involves an injustice from the perspective of corrective justice, since D has no duty to pay compensation, but it also achieves some retributive justice – D suffers a harm (paying damages D does not owe) in response to the culpable conduct. A false negative, by contrast, involves a failure of both corrective *and* retributive justice: C does not receive his entitlement in corrective justice and nor is D punished. But if we really think that D deserves to be punished, why leave the punishment to chance? If we think that D merits punishment, then D should be punished – perhaps by awarding C exemplary damages. There still seems to be no explanation why punitive considerations should take effect at the evidential level.

Responsibility

There may be conditions under which the defendant can meaningfully be said to be responsible for the causal uncertainty faced by the claimant.[71] This might justify some form of alleviation of the orthodox causal proof requirement. This idea has been recognised in English law. For example, in *Shawe-Lincoln* v *Neelakandan*, the causal issue was whether the defendant's negligent failure timeously to refer the elderly claimant to hospital was a cause of (part of) his injury.[72] Lloyd-Jones J (as he then was) held that the fact that the defendant GP deprived the claimant of the opportunity to have tests in hospital which would have revealed crucial details about the state of his deteriorating health justified a 'benevolent' approach to assessing the claimant's case on causation.[73] The fact that such tests were not available rendered the claimant's case on causation harder to prove.

If the defendant is in some way responsible for the causal uncertainty faced by the claimant, this intuitively provides a reason for alleviating the claimant's burden of proof on causation.[74] This is so even where the defendant's responsibility for the uncertainty does not itself have some probative force, as where the defendant intentionally destroys some medical records in anticipation of litigation, leading to an inference that the evidence would have been damaging to the defendant. Rather, the defendant's responsibility for the uncertainty faced by the claimant seems to justify some alleviation in the claimant's burden of proof independently of the epistemic significance of the defendant's conduct. But I will argue here that a plausible account of the conditions of such responsibility will be a restrictive one.

It will be helpful to start with the most developed account of this responsibility, given by Porat and Stein.[75] The defects in this account, I will suggest, point towards a narrow doctrine. Porat and Stein argue that some legal response

71 See generally Porat and Stein, n 18 above.

72 [2012] EWHC 1150.

73 *ibid* at [81]. The claimant nonetheless failed on the causal issue since, even allowing for a benevolent approach, the evidence still pointed against the conclusion that the defendant's negligence was causative.

74 Or, alternatively, recognising a claim in respect of the damaged evidential position itself. For an explanation of the relative merits of these alternatives, see Porat and Stein, n 18 above, 165–169.

75 Porat and Stein, *ibid*.

can be justified in situations where D's wrongful conduct was a cause of C's foreseeably being in an evidentially worse position in relation to proving C's case on causation against D.[76] In such circumstances, D has inflicted what they call 'evidential damage' upon C.

Their account is particularly broad because of the range of situations in which they claim the defendant has inflicted evidential damage. Though they do not explicitly say so, it seems that the 'evidential damage' is inflicted in every case in which the claimant, without fault in relation to the absence of evidence, finds itself in a position where it cannot prove on the balance of probability that the defendant's wrongful conduct was a cause of its injury, due to the defendant's negligence.[77]

So, for example, they claim that the defendant doctor inflicted evidential damage in *Hotson* v *East Berkshire AHA*[78] (*Hotson*). In that case, the defendant negligently failed to diagnose the claimant when he first presented at hospital. The claimant then suffered avascular necrosis. There was a 25 per cent chance that C's avascular necrosis would have been prevented by non-negligent, timeous, diagnosis and treatment. The claim failed since it was more likely than not that the injury would have occurred in any event. Since *Hotson* is structurally the same as virtually any case where there is uncertainty over whether wrongful conduct made a difference to the claimant's situation, the evidential damage doctrine has very wide scope.

These claims are conceptually flawed. Firstly, in order for a person to be evidentially damaged, it must be that there would be more evidence or information in the hypothetical world where the defendant did not behave wrongfully as to whether the defendant's wrongful conduct was causative than in the actual world.[79] In cases where it cannot be established on the balance of probability that the defendant's wrongful conduct was a cause of the injury in the actual world, this will frequently not be the case. For instance, in *Sienkiewicz*, had the defendant not behaved negligently, it may have been the case that the injury would not have occurred.[80] In which case, it would be meaningless to say that there would have been more evidence about the negligence's causative effect.

Even where, on the balance of probability, the injury would still have occurred had the defendant not behaved wrongfully, the only new information

76 Their preferred response in most cases is a claim for damages in respect of the 'evidential damage' itself, where the damages would be calculated in proportion to the probability that the evidence would have favoured the claimant's case, but they acknowledge that responsibility for evidential damage can justify alleviation of the general proof rules in some cases. See Porat and Stein, *ibid*, 166.

77 For a similar verdict, see R. W. Wright, 'Liability for Possible Wrongs: Causation, Statistical Probability, and the Burden of Proof' (2008) 41 *Loyola of Los Angeles Law Review* 1295, 1306 n 43: '. . . applicable to any case in which there is any uncertainty [over causation]'.

78 n 27 above.

79 Their statements typically suggest that *in the absence of the defendant's wrongful conduct, there would be information* surrounding the cause of the injury: 'By inflicting evidential damage and thereby taking away from the plaintiff the information necessary for ascertaining the cause of her or his direct damage, the defendant took something of value from the plaintiff', Porat and Stein, n 18 above, 167. cf *ibid*, 201.

80 n 5 above.

which would have been in existence in, say, *Hotson*, is the eliminative evidence that the defendant's wrongful conduct could be ruled out as a cause of the claimant's injury. If the defendant did not behave wrongfully, it could not have been a cause of the injury which occurred. But it is not plausible to consider this fact to constitute damage or loss to the claimant. What is the value of the information that wrongful conduct could be ruled out as a cause of the claimant's injury in the event that the defendant did not behave wrongfully? Perhaps it provides some peace of mind – 'I am injured but at least not through wrongful conduct [!]' – and might assist in the decision whether or not to sue. But this weak interest could hardly justify either substantial damages or a significant alleviation in the claimant's burden of proof.

Secondly, if the justification for the evidential damage doctrine is that individuals have a valuable interest in having evidence concerning their legal claims, it follows that this interest must exist prior to the defendant's interfering with it. But in most cases where the defendant's wrongful conduct is alleged to be both a cause of the primary injury *and* the absence of information about the cause of that injury, the same act which gives rise to the interference with this valuable interest also *ipso facto* creates this valuable interest. Without the wrongful conduct, there would be no interest in having evidence concerning that wrongful conduct. This fact makes it difficult to say that the interest has been set back or damaged. In the same way, when a builder builds a defective house, she has not 'damaged' the house.[81]

There are also normative problems with Porat and Stein's claims. First, the proposal is inconsistent with the normative foundations of the burden of proof on causation. It follows from their view that each of us should have legal duties not to act negligently (or more generally, wrongfully) in such a way that one's negligence cannot be ruled out as a cause of another's injury. This duty is breached in every case where it is foreseeably uncertain whether one's negligence was a cause of the claimant's injury. But, if so, what room is left for the general proposition that it is for the claimant to establish their case on causation? Second, the justification for this duty, on their view, is an autonomy-based interest in having information relevant to one's legal claims.[82] If this is right, it is difficult to believe that the only requirement for the existence of the duty should be reasonable foreseeability of interference with another's autonomy in relation to their legal claims. If liability for foreseeably causing thousands of pounds of (purely) economic loss depends at least upon an assumption of responsibility between the parties, liability for deprivation of information relevant to a decision about whether to bring a legal claim should require at least the same.

Once the conceptual and normative problems in Porat and Stein's account are remedied, a considerably narrower doctrine emerges. To remedy the conceptual problem, it is necessary to insist either that (a) the person who inflicts the evidential damage is different from the person whose wrongful conduct is alleged to have caused the primary damage or (b) that defendant's conduct can be said to have interfered with an evidential interest which existed independently of at

81 *Murphy* v *Brentwood District Council* [1991] AC 398.
82 Porat and Stein, n 18 above, 177.

least a part of the defendant's conduct. This itself assists in remedying the first normative problem, since it will not be every case where (a) or (b) is satisfied. The second normative problem can be met by insisting at least upon a special relationship between the parties whereby one person assumes responsibility for another's autonomy interests.

What situations, then, would fall within this narrower doctrine? It is certainly a requirement of doctors' duties of care, in certain circumstances, that they maintain information. A doctor's duty to take care of a person's health will entail that it must keep good records for future treatment – lest the doctor endanger the person's health. But we should also say that doctors owe a duty to account for the course of the treatment, where the normative reasons for that duty extend beyond preservation of the patient's health. As Porat and Stein argue: 'this duty can and, indeed, should be derived from the patient's general right to receive full information concerning her or his diagnosis and treatment'.[83] For instance, a surgeon would breach this duty by carelessly failing to maintain records of an operation which went badly or by disposing of a swab used during such an operation, where that would likely provide evidence as to the cause of the patient's injuries.

If the conditions of this narrower responsibility are met – a genuine deprivation of evidence and an assumption of responsibility for a person's autonomy interest – the issue arises of how the law should respond in relation to the claimant's burden of proof. The answer lies in identifying correctly how the defendant's conduct has affected the claimant's evidential position. Of course, if the court knows exactly how the missing evidence would have affected the claimant's case, then, in effect, the claimant has suffered no evidential damage. But in most cases the effect of the defendant's deprivation will be either (a) to undermine the court's confidence in the assessment of the probability of the claimant's case on causation, because there is less evidence on which to base that assessment and/or (b) to deprive the claimant of evidence which could have altered the probability of its case on causation, but the extent to which it would alter the probability is unknown. For instance, if the claimant can only establish its case on causation with fairly general statistical evidence due to the defendant doctor's failure to maintain any records, but can 'prove' its case on the balance of probability using this evidence, then it should succeed. Because the defendant's wrongful conduct has deprived the claimant of the ability to establish its case using case-specific evidence, it is fair that it is not held to this standard in proving its case. If (b) is true, English law's current approach seems reasonable: the court should be more willing to find for the claimant on a factual issue to which the missing evidence relates.[84] This is the meaning of the 'benevolent' approach.[85] So a missing set of test results should be presumed to bear the most favourable results consistent with the remaining body of evidence. As the

83 Porat and Stein, *ibid*, 188. It could also be supported by the effect that failing to keep records may have on the ability to ability to enforce secondary rights in respect of physical injuries, given the argument above, 12–13.

84 See above, 749.

85 See n 73 above, 749.

Neelakandan case demonstrates, it need not follow that the claimant will then succeed on the causal issue, since the missing evidence may only relate to one strand in the claimant's case on causation.[86] A favourable presumption on that strand need not entail overall success on the causal issue in the face of other, conflicting, evidence .

Incentives-based

Suppose that in a particular context potential injurers can readily predict *ex ante* that it will be impossible *ex post* for a person potentially injured by their conduct to prove that that conduct was a cause of injury suffered. If such circumstances exist, it may be that potential injurers have less incentive to conform to their legal duties in those circumstances. Altering the burden or standard of proof might resurrect the incentives blunted by the existence of predictable causal uncertainty. To remedy these weakened incentives, one might (a) reverse the burden of proof, (b) reduce the standard of proof or (c) establish liability in proportion to the probability of causation. Which remedy one should select would depend on the reason proof is impossible in the particular context. If proof is impossible because there is a recurrent statistical probability of causation below 50 per cent, then (c) would be the optimal solution, since (a) would allow the defendant to escape liability by pointing to the probability of causation, whilst (b) would not cover cases below the standard of proof. By contrast, if proof is impossible because there is simply no available way of quantifying the probability of causation, then (a) would be the optimal solution. Call this line of argument the 'incentives argument'.[87]

This argument was made in *McGhee* v *National Coal Board* to justify accepting proof that the defendant's negligence materially increased the risk of the claimant's dermatitis as sufficient in lieu of proof of causation on the balance of probability, in circumstances where the non-tortious risk of injury imposed by the defendant's activities, and absence of scientific understanding over the mechanism by which dermatitis is caused by exposure to brick dust, precluded such proof.[88] This was the argument that the defendant's duty of care could be ignored 'with impunity', since it could never be established, in such circumstances, that the breach of the duty was a cause of the claimant's injury on the balance of probability.[89]

86 *ibid*.

87 For a clear statement, see A. Porat, 'Misalignments in Tort Law' (2011) 121 *Yale Law Journal* 82, 108–114.

88 *McGhee* v *National Coal Board* n 14 above.

89 *ibid*, 6 *per* Lord Simon. The argument – rephrased as the idea that the duty would be 'empty of content' – was then relied upon subsequently in *Fairchild* v *Glenhaven Funeral Services Ltd* n 53 above at [62] *per* Lord Hoffmann. An incentives-based argument has also been relied upon to justify the rule that a fiduciary cannot plead that its principal would have consented to its breach of fiduciary duty in order to avoid liability to disgorge gains made in breach of that duty: *Murad* v *Al-Saraj* [2005] EWCA Civ 959 at [33].

The argument has also been accepted, more generally, in German law in certain cases involving contractual duties of disclosure, where the defendant in breach of contract must prove that, if informed, the claimant would have decided in the same way.[90] Echoing the English law reasoning here, it has been stated that duties of disclosure would be emptied of content if the claimant was put to proof that it would have acted differently if properly informed.[91] Such proof, it is said, can rarely be brought.[92]

The incentives argument requires empirical (will the rule proposed by the argument in fact lead to better outcomes (fewer losses or wrongs)?) and normative assessment (assuming the rule generates the consequences expected, is there nonetheless some moral constraint which requires us not to produce these consequences?).

In the absence of any direct empirical study comparing the effects of a liability system which uses one of (a)–(c) in situations of impossibility of proof and a system which does not, the only way to proceed to answer the empirical question is to adopt some model of the effects of liability.[93] The model assumed by the proponents of the incentives argument assumes that agents act in order to maximise their own expected utility (or wealth) and typically that agents face no incentives to conform to their tortious obligations other than the threat of sanctions provided by tort law.[94]

If agents behaved in this way – as individual expected-utility-maximisers – and faced no incentives other than those provided by the threat of compensatory remedies in tort law, *and* could predict that causal proof would be impossible in a particular situation, then we would expect more injuries to occur without an exceptional rule being adopted. This is because such agents could save on the costs of precautions without facing any adverse consequences.[95]

We can have little confidence in this model, however, since it involves multiple assumptions which diverge considerably from reality. Firstly, there is the standard, tired-but-true, criticism that individual agents are not individual expected utility maximisers. This general criticism is likely to have significant consequences for the analysis of the specific situation of causal uncertainty.

One reason is that the notion of impossibility of proof is itself probabilistic. An agent will almost never be 100 per cent certain that it will be impossible to prove causation against the agent in a subsequent trial. Rather, the agent will have some

90 BGHZ 61, 118. See also *Chester* v *Afshar* [2004] UKHL 41 at [87] *per* Lord Hope.

91 BGHZ 61, 118, 121.

92 *ibid*.

93 Some effects of the adoption of probabilistic liability rules, such as the effect on the number of claims and on insurance premiums, have been studied: see S. Koch, 'Whose Loss is it Anyway – Effects of the Lost Chance Doctrine on Civil Litigation and Medical Malpractice Insurance' (2010) 88 *North Carolina Law Review* 595, 619–631 (detecting little effect).

94 eg, A. Stremitzer, 'Negligence-based Proportional Liability' 34–37 at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2088977 (last accessed 22 May 2015).

95 *ibid*, 37.

(more or less accurate) subjective probability that this is so.[96] Much will depend, then, on how individual agents make and act upon these probability assessments. If agents are afflicted with a pessimistic bias, they may place great weight on even a small probability that proof will turn out to be possible, with the result that efficient care will still be taken.[97] Alternatively, if agents have an optimistic bias, they might overestimate the probability that taking precautions will lead to an increased probability of harm being avoided.[98] Consequently, they may overestimate the *ex post* probability that their conduct would have prevented the harm and thus decide to take efficient care.

Second, it is quite surprising that even (in other respects) sophisticated economic analyses assume that the only relevant incentives in play are those provided by compensatory remedies in tort law.[99] This is almost never true. Morality, market forces, the criminal law, social norms, and regulatory sanctions may exercise some effect on behaviour. It is difficult to have much confidence in a model of the behaviour of medical professionals who deal with the most vulnerable individuals who have placed trust in those professionals, which does not allow for the (significant?) possibility of some moral concern influencing behaviour.[100] Even tort law may provide incentives simply by declaring doing-x to be a legal duty without sanctions being available for the failure to do x.[101] Some people might simply think it is important to comply with the law for its own sake.

In short, we can, at present, have little confidence in existing models of the effects of orthodox liability rules in situations of recurrent causal uncertainty. The next question is whether, given the little that is currently known, it would be justifiable, in relation to particular defendants, to impose liability without proof of causation in contexts where we believe that proof is most often predictably impossible with some degree of confidence, on the ground that this *possibly* would lead to a reduction in the overall number of injuries. If we think that causation is normatively essential to grounding the defendant's moral liability to pay compensation to the claimant, and we think that individuals have a moral right to be free of state coercion unless a case is proven against them, it would be difficult to accept that the speculative possibility of some injuries being prevented would be enough for us to violate that right. These are large 'ifs',

96 The point here is not that the mere fact whether proof is impossible is probabilistic is an objection. It would still theoretically lead to underdeterrence if there were only, say, a 40% probability that proof would be impossible. The point is that how these probabilities are acted upon may well be affected by cognitive bias.

97 On pessimism as to risk more generally, see J. Teitelbaum, 'A Unilateral Accident Model under Ambiguity' (2007) 36 *Journal of Legal Studies* 431, 433.

98 On optimism as to risk more generally, see E. Posner, 'Probability Errors: Some Positive and Normative Implications for Tort and Contract Law' (2003) 11 *Supreme Court Economic Review* 125.

99 Stremitzer, n 94 above, 37: 'Imagine a world without liability. An injurer who may inflict harm on others knows that he will never bear the consequences of his actions and therefore has no incentive to invest in safety in order to prevent accidents'.

100 The point has not been ignored by legal economists, but it does not enter into standard models of accidents. See, eg, L. Kaplow and S. Shavell, 'Moral Rules, the Moral Sentiments, and Behavior: Towards an Optimal Moral System' (2007) 115 *Journal of Political Economy* 494, 495.

101 S. Smith, 'The Normativity of Private Law' (2011) 31 OJLS 215.

which I have done nothing to support here, but they at least have some prima facie plausibility.[102] If the current state of our knowledge is as I describe it, the more difficult question of what we should do where we do have clear, credible, evidence that departing from orthodox requirements of proof of causation in tort law would prevent injuries, and that such effects could not better be achieved by, for instance, regulation, can be postponed.

Relative cost of errors

When we decide upon what proof rules we should have on causation, it is rational to consider the relative costs of false positives and false negatives. If a false positive (finding D to have caused C's injury when this is not so) is much worse (in some sense) than a false negative (finding D not to have caused C's injury when this is so), then this would suggest that we should have a proof rule which favoured the latter.[103] The balance of probability rule displays a slight preference for false negatives: if the parties' claims are equally balanced, the claimant loses.

But it could be argued that it is irrational to ignore the fact that a false negative could have a crushing effect on an individual but minimal effect on a large corporation. If so, and especially if and where the substantive law is grounded in such considerations of fair loss distribution, it may be argued that the orthodox proof of causation rule is not justified, since it irrationally privileges false negatives over false positives. In short, one possible claim is that the greater capacity to bear a loss (where 'capacity' relates to the effect on each individual's welfare) between the parties should be considered in weighting false positives and negatives. In distributing the risks of uncertainty, the effects of errors on the parties should be taken into account. A related argument might be that the effects on distributive justice should be taken into account in the weighting: suppose that a false positive will only have the effect of reducing the defendant's unjust holdings, whereas a false negative will have regressive effects on the claimant's position in distributive justice. Finally, one might simply hold that certain kinds of error in the determination of liability are not as serious as others. For example, perhaps a false positive on causation against a public authority in a claim by an individual against that authority is not as regrettable as false negative in favour of the authority against the individual.

Clearly, these arguments raise very large questions. The extent to which it is legitimate to rely upon such arguments in tort law is disputed.[104] Here I will simply make two observations. First, if such arguments could infiltrate the inquiry into how false positives and false negatives should be weighted, there would be no clear reason to restrict them to determinations of causation. Yet it seems hard to accept that a defendant's claim to have fault proven against them

102 For the outlines of a defence, see B. Zipursky and A. Ripstein, 'Corrective Justice in an Age of Mass Torts' in G. Postema (ed), *Philosophy and the Law of Torts* (Cambridge; CUP, 2001) 221–225.

103 There is a substantial decision-theoretic literature which builds on this simple point. See, eg, J. Kaplan, 'Decision Theory and the Factfinding Process' (1968) 20 *Stanford Law Review* 1065.

104 See generally on distributive justice arguments, T. Keren-Paz, *Torts, Egalitarianism and Distributive Justice* (Aldershot: Ashgate, 2007).

(in fault-based causes of action) should be diminished by, say, their (much) greater comparative wealth.

Second, the greater ability of one party to bear the loss is really independent of the truth or falsehood of the case on causation. It is not that an *incorrect* verdict harms the claimant more than the defendant. Suppose that the defendant did not cause the claimant's injury. It would still be true that the defendant could better bear the loss the claimant; it would still be the case that not awarding the claimant damages would harm the claimant more than awarding damages would have the defendant (in terms of the effect on the welfare of each).

CONCLUSION

This article has defended the justifiability of making exceptions to the basic rule of proof of individual causation. This sounds radical, but the argument has in fact been a largely conservative one. Under none of the rules defended – the causative defendant indeterminacy rule, the claimant indeterminacy rule, or the defendant indeterminacy rule – are defendants held liable when they have not been proved to have caused injury.[105] Proof of causation is preserved, even if proof of individual causation is not. In this way, the basic moral concern to which exceptional rules give rise – that such rules undermine a commitment to individual responsibility – is assuaged.

If the arguments of this article are correct, then the factual causation inquiry in tort law can be approached by first asking – Was D's wrongful conduct a but-for or material contribution cause of C's injury on the balance of probabilities? If it was not, a secondary question arises: do any of the following rules apply in relation to D: (i) the causative defendant indeterminacy rule, (ii) the claimant indeterminacy rule, or (iii) the defendant indeterminacy rule?

Here is one short-cut heuristic for determining whether any of (i)–(iii) apply. One asks: Would the effect of applying the normal rule in relation either to a particular claim before the court or to a set of claims before the court be, on the balance of probabilities, to deprive a person of a right to compensation in respect of an injury which would not have occurred had no one behaved wrongfully in relation to that person? If the answer is positive, then one of (i)–(iii) will apply on the facts.

However, recourse to the individual sub-rules in (i)–(iii) remains necessary for two reasons. First, it will be relevant in determining the *extent* of liability: if D is liable under (i) or (ii), care needs to be taken that D is not held liable for an amount of loss which it is known not to have caused. Second, the justificatory case for each rule is different; (i) and (ii) are relatively straightforward to justify,

105 It might be claimed that the logical conclusion to be drawn is that the rules defended should not be considered 'exceptional' at all. If the general rule is defined, as is typical (see n 5 above), as requiring that a claimant prove that the defendant (from whom it seeks to recover damages) caused its injury, then the CDIR and CIR *are* exceptional: they permit the claimant to obtain damages without such proof. The *Defendant indeterminacy rule* could be classified as simply a complex application of the general rule. However, it is preferable to consider it to be a distinctive rule, since (a) it does involve a non-standard equiparation of primary and secondary rights and (b) it permits liability even where the *prevention argument* does not operate (see n 107 below).

whilst (iii) is not. This is because under (i) and (ii) the defendant is simply being required to pay damages in respect of a straightforwardly actionable injury that it has caused, albeit that the identity of its victim cannot be established. Where all defendants are solvent, each defendant's liability under (iii) is also in respect of injury caused,[106] but the argument relies upon treating the interference with bodily integrity and interference with secondary rights in relation to bodily integrity as sufficiently similar kinds of injury.[107] These differences could be relevant in the event that some extension to the rules is envisaged.

106 Albeit that *which* defendant caused *which* injury is not established.

107 If not all defendants are solvent in a defendant indeterminacy situation, then it is necessary to rely upon the other justifications for liability given above at 741–742.