FACTUAL CAUSATION

Jane Stapleton*

The doctrinal parameters of the tort of negligence are remarkably open-textured which is why it has typically been in negligence cases that foundational formulations of factual causation have been made. This area of law has recently undergone an extensive restatement by the American Law Institute ('ALI') and been the subject of legislative attention in all Australian states. In the light of these developments this essay sketches some essential issues relevant to factual causation which apply not only to the tort of negligence but throughout the law.

I THE RESTATEMENT OF US TORT LAW

At the end of the 20th century the ALI launched a project to compile a Third Restatement of the general principles of the US law of torts. Previous restatements had covered a field described as 'legal causation' so the question arose as to how to restate this field. The orthodox starting point for common law analysis is to read the cases and deduce there from the meaning courts ascribe to relevant terms such as 'duty' or 'breach'. This most basic of analytical techniques fails utterly in the area of causal terms. In the First and Second Restatement of Torts the term 'legal cause' signified an amalgam concept consisting of both an historical connection element (did the breach contribute to the injury?) and a truncation of legal responsibility element (should the party in breach be liable for this injurious consequence of the breach?). Yet in some case law 'legal cause' signified only the truncation of legal responsibility element. Exactly the same terminological disarray was present in relation to the term 'proximate cause': in some contexts it signified the amalgam of the two elements; in others, the most common usage, it referred only to the truncation issue. So it was that US case law contained both: statements that proximate cause was the second component of legal cause;\(^1\) and statements that legal cause was the second component of proximate cause.\(^2\)

Compounding this confusion was the indiscriminate deployment by US courts of the term 'substantial factor'\(^3\) (akin to the shifting use of 'material contribution' in the Commonwealth).\(^4\) Sometimes 'substantial factor' was used in relation to the

---

* Professor, Australian National University.

1 See, eg, Winn v Posadas, 913 A 2d 407, 411 (Conn, 2007).

2 See, eg, Harrison v Binnion, 214 P 3d 631, 638 (Idaho, 2009).

3 See, eg, Dan Dobbs, The Law of Torts (2000) 416: 'the substantial factor test is not so much a test as an incantation'.

requirement that the legal complaint relate to an injury that is more than trivial. Elsewhere it was used: as a synonym for the 'but-for' test; or as a fudge to mask the inadequacy of the but-for test of historical connection in cases such as where there are two or more sufficient factors;\(^5\) or to mask radical rules developed to permit a plaintiff to jump an otherwise unbridgeable evidentiary gap.\(^6\)

There was no value in the Institute attempting to 'restate' this causal usage. Clarity of exposition required that causal terminology be confined to a single idea, namely the objective idea of historical connection, conveniently captured by the term 'factual causation'. It was crucial that this question of fact (and the special rules of law relating to its proof) be recognised as completely distinct from the issue of where and why responsibility for the infinite chain of consequences of conduct should be truncated. This truncation issue, sometimes known outside the US as 'remoteness', rests entirely on the \textit{normative} analysis of the facts. Accordingly it was recommended to the ALI:\(^7\) that each of these two issues be given its own separate chapter in the new Restatement; that terms such as 'legal cause', 'proximate cause' and 'substantial factor' should be completely abandoned; and that henceforth, the truncation issue should be described in the completely non-causal terms of 'the scope of liability for the consequences of breach', or 'scope of liability' for short.

It was a reflection of the deep dissatisfaction with the state of US doctrine in the area, and in particular with the description of the truncation issue in causal terms, that all these suggestions were smoothly adopted by the ALI, despite their radical nature both in terms of their departure from the terminology of case law and in terms of the structure of preceding restatements of torts. Accordingly in volume 1 of the \textit{Restatement Third, Torts: Liability for Physical and Emotional Harm} (2009) Chapter 5 deals with the issue of 'Factual Cause' while a completely separate chapter, Chapter 6, deals with the distinct truncation issue of 'Scope of Liability'.

\section{THE COMMONWEALTH}

What then about the state of case law on these issues in the Commonwealth? Remarkably, here there has been even more disarray for, whereas US courts had long accepted that there were two separate issues at stake, the historical connection issue and the truncation issue, Commonwealth courts have struggled to express this separation in a consistent and coherent manner. One major source of difficulty in Australian courts has been the frequent and lamentable recourse to the slogan of 'common sense' causation.\(^8\)

---

5. See, eg, \textit{Anderson v Minneapolis, St Paul & Sault Ste Marie Railway Co}, 146 Minn 430 (Minn, 1920) a case involving the merging of two fires where the term was first coined by a US court.


Of course in some cases what the court means by this slippery term is simply the permission to infer facts from common experience. Even here, however, appeals to ‘common sense’ can mislead. For example in *Naxakis v Western General Hospital* Justice Gaudron stated that:

For the purpose of assigning legal responsibility, philosophical and scientific notions are put aside ... in favour of a common sense approach which allows that ‘breach of duty coupled with an [event] of the kind that might thereby be caused is enough to justify an inference, in the absence of any sufficient reason to the contrary, that in fact the [event] did occur owing to the act or omission amounting to the breach’.9 (citing *Betts v Whittingstowe* (1945) per Dixon J).10

Recent judgments in the High Court such as *Roads and Traffic Authority v Royal*11 helpfully highlight the danger in such ‘common sense’: namely that a court will elide proof of breach (which increased the risk of a certain outcome) and foreseeable result (of the same type of outcome) with proof that the breach was a factual cause of that result in the specific case at hand, and fail adequately to consider whether there was, on the evidence, ‘any sufficient reason to the contrary’. In *Betts* Dixon J was himself careful to support the inference he drew with an explicit finding that ‘the facts warrant no other inference inconsistent with liability on the part of the defendant’.12

But more often the term ‘common sense causation’ has been used by Australian courts in an attempt to navigate the analytical morass that results from a failure clearly to distinguish historical involvement and truncation. For decades the result has been that across swathes of Australian case law the deployment of causal terminology has been muddled and often incomprehensible, obscuring the underlying reasoning of the court. It brings the law into disrepute if, when confronted with a hotly disputed complex dispute about the appropriate point at which legal liability should be truncated, a court accepts the ‘glib submission’13 that its resolution rests on nothing much more than ‘common sense’.

Resort to this problematic device was greatly fuelled by the publication in 1959 of *Causation in the Law* in which Herbert Hart and Tony Honoré sought to study the usage of causal terms by lawyers which they argued had tracked ‘the plain man’s use of causal notions’.14 Importantly, while the book was being written Hart spent a sabbatical leave at Harvard where he observed at close quarters US lawyers using causal terms such as ‘proximate cause’ and ‘legal cause’ to refer to or at least encompass the truncation issue. Influenced by such usage, Hart and Honoré asserted that both the question of historical contribution and the question of truncation of responsibility were ‘causal’ questions: that causation in the law was ‘bifurcated’.

From selected data of causal usage, Hart and Honoré extracted what they called ‘common sense principles of causation’. This is not the place to rehearse all the grave

---

9 (1999) 179 CLR 269, 281 [36], quoting *Betts v Whittingstowe* (1945) 71 CLR 637, 649 (Dixon J) (‘Betts’).
10 (1945) 71 CLR 637, 649 (Dixon J).
12 *Betts* (1945) 71 CLR 637, 649.
difficulties in their complex approach. But it is worth looking at one of their elaborate and ultimately unhelpful 'principles' because it illustrates the sort of textual obfuscation that runs through key Australian decisions. In looking at cases which turned on the truncation issue they found that it was often the case that liability was denied when, after the defendant's breach of obligation, a third party or an abnormal event (such as lightning) had intervened. Australian courts often communicated this result by stating that the intervention 'broke the chain of causation' while, as we have seen, American courts typically did so by stating that the tort was not the 'legal cause' or was not the 'proximate cause' of the outcome. Since Hart and Honoré were committed to the project of mapping causal usage they concluded that:

courts have often applied, in their determinations of causal questions, a central concept in which great emphasis is laid on voluntary action or abnormal and coincidental events as negating causal connection.

But the fact remained that there were myriad situations where the law imposed liability in the presence of just such an intervention. While the fact of intervention of a third party or abnormal event certainly may be relevant to our enquiry about where liability should be truncated, it is not clear from Hart and Honoré when and why the intervention might be relevant: what they produced was merely a topology of causal usage, not a geology of the normative reasoning lying beneath that usage. This absence of normative rationale exposes a judge to the temptation of merely asserting a conclusion on the truncation issue without providing reasons, while reciting some version of Hart and Honoré's 'central concept' of causal connection — a temptation to which Justice McHugh fell victim on a number of occasions. This is hardly an advance on what Hart and Honoré rightly called the 'obscure metaphor' of the intervention 'breaking the chain of causation'.

The temptation to mere assertion was further exacerbated by Hart and Honoré's characterisation of their observed patterns of truncation in terms of causal connections. Just as the Americans found that giving the truncation issue causal names such as 'proximate cause' or 'legal cause' carried with it the risks of confusing jurors, jury-free jurisdictions also face risks if the truncation issue is characterised as a 'causal' question. This is because for many of us the notion of causation has a factual ring. In ordinary speech we tend to think of something either being a cause or not, and we often do not see our conclusions on the matter as requiring normative justification. In short, so long as the truncation issue is framed in causal language, some judges will be tempted to present their determinations relating to truncation without adequate normative

---

15 See Jane Stapleton, 'Choosing What We Mean by "Causation" in the Law' (2008) 73 Missouri Law Review 433, 458–65 (hereafter 'Choosing').

16 Hart and Honoré, above n 14, 130–1.

17 For example sometimes there is liability even though lightning has intervened and sometimes there is no liability: we need to delve into the reasons why this is so, see Stapleton, 'Choosing', above n 15, 461–4.


19 Hart and Honoré, above n 14, xxxii.

20 Indeed, Hart and Honoré even characterised their truncation notions as not just causal but 'factual': Hart and Honoré, above n 14, lli, 91.
justification. Accompanying this is the risk that, so long as both the factual issue of historical involvement and the normative issue of the truncation of liability are framed as 'causal' questions, a trial judge may not easily recognise whether statements in previous appellate cases concerning ‘causation’ relate to historical involvement or truncation. Judges would better communicate their reasons if the historical contribution question was kept absolutely separate from the truncation question and if causal terminology was confined to the former.

The argument for such clarification of terminology\(^{21}\) was referred to with approval in the final report of the Review of the Law of Negligence, the ‘Ipp Report’ which recommended the clear separation of factual causation and the scope of liability.\(^{22}\) Bizarrely, however, the Ipp Report chose to retain the umbrella term of ‘causation’ to signify the amalgam of both issues. So it was that, at the very time the American Law Institute was stripping the truncation issue of its misleading causal label throughout the law of torts, the Ipp Report was entrenching that barrier to clarity of legal analysis in Australia: for all Australian States plus the Australian Capital Territory\(^{23}\) followed the Ipp Report recommendation and legislatively adopted the ‘causation’ umbrella term albeit only in the limited field covered by that Report, namely where the focus is on the 'fault of a person (the "tortfeasor"),'\(^{24}\) 'negligence'\(^{25}\) or a 'breach of duty.'\(^{26}\) So long as the 'causation' term is used as such an umbrella concept, it renders incoherent the many judicial statements that 'causation is essentially a question of fact.'\(^{27}\)

Nothing would be lost and much would be gained if Australian courts quietly ignored the umbrella term both under these civil liability statutes where it does no substantive work and, elsewhere, resisted the temptation to refer to 'common sense causation' and proceed directly to the analysis of the separate issues of factual cause and scope of liability. This substantial improvement in the clarity of exposition of judicial reasoning would, it is to be hoped, then work its way into legislative drafting which has been so bedevilled by cryptic terms aimed at capturing some amalgam of factual cause and scope of liability.\(^{28}\)

21 See Stapleton, ‘Cause-in-Fact’, above n4
23 Civil Liability Act 2003 (Qld) s 11; Civil Liability Act 2002 (NSW) s 5D; Wrongs Act 1958 (Vic) s 51; Civil Law (Wrongs) Act 2002 (ACT) s 45; Civil Liability Act 2002 (Tas) s 13; Civil Liability Act 1936 (SA) s 34; Civil Liability Act 2002 (WA) s 3C. Neither the Northern Territory nor the Commonwealth has implemented any provision relating to 'causation'.
24 Civil Liability Act 2002 (WA) s 5C(1).
25 Civil Liability Act 2002 (NSW) s 5D(1); Wrongs Act 1958 (Vic) s 51(1); Civil Law (Wrongs) Act 2002 (ACT) s 45(1); Civil Liability Act 1936 (SA) s 34(1).
26 Civil Liability Act 2003 (Qld) s 11(1); Civil Liability Act 2002 (Tas) s 13(1).
28 All liabilities, including those arising under statute, are limited. A statute may expressly limit the type of consequence that comes within its scope; see eg, Allianz Australia Insurance Ltd v GSF Australia Pty Ltd (2005) 221 CLR 568. Or such limits may be generated implicitly by the clear purpose of the statute; eg, Gorris v Scott (1874) 9 LR Exch 125. More often these must be divined by the court from more general interpretations of the purpose of the rule in the light of the general law.
III DOCTRINAL REQUIREMENTS ANALYTICALLY PRIOR TO AND AFFECTING THE FACTUAL CAUSE ISSUE

Factual cause is not the only element of any plaintiff’s cause of action. For example, in the tort of negligence a plaintiff must show her claim to be a type of complaint that is actionable in this tort: she cannot, for example, complain of discomfort from noise or smells, even though these may be actionable in another cause of action such as the tort of private nuisance. The list of types of complaint that are actionable in a cause of action is not closed. One important area of current controversy is whether a negligence plaintiff will be allowed to complain about the loss of a chance. This is relevant to my topic tangentially because sometimes courts create a special rule to assist a plaintiff and it is not clear whether the rule extends the list of complaints that are actionable or is a special rule about proof of factual cause.

Next, the plaintiff must establish that the defendant owed her a duty of care in relation to the damage of which she complains. Traditionally, duty operates as an ‘incidence’ rule by specifying the contexts in which, were a defendant to have appropriate legal advice before he conducted himself, he could be advised that he owed ‘such and such’ a legal obligation. For example in relation to physical injury we owe a duty to the whole world when we engage in affirmative action. The nature of that duty to the whole world is always the same: simply to take care to avoid causing physical injury to anyone (though its content, that is what reasonable care requires in the circumstances, will vary).29

One could, of course collapse all the elements of the tort of negligence — actionable damage, duty, breach, factual cause and scope — into one amorphous unstructured proposition of law: that in the circumstances the defendant owed this plaintiff a duty of reasonable care not to cause this particular outcome of which the plaintiff complains. But this merges, and therefore dulls to the point of uselessness, any specific messages about, and themes within, the underlying legal concerns influencing the outcome of the case. For example, the duty analysis provides the opportunity to signal systemic concerns such as the concern with individual autonomy that underlies the no-duty-to-rescue-a-stranger doctrine. Moreover, it does not help people understand, and to the extent possible, be guided by the law.30

Alas, this was the step Lord Hoffmann took in South Australia Asset Management Corp v York Montague Ltd.31 In communicating his decision that a valuer was not liable for a particular outcome suffered by the plaintiff, his Lordship asserted that such an outcome did not fall within ‘the scope of the duty’ of the valuer. In making this move His Lordship not only collapsed the specific truncation issue in the case into some duty-breach amalgam but also failed to explain and justify why liability of this valuer did not extend to the particular loss.

29 The situation is more complex where the duty is one of affirmative action or the form of actionable damage in issue is nervous shock or pure economic loss. Here there is a normative ‘envelope’ confining the obligation to certain types of risk; see Jane Stapleton, ‘The Risk Architecture of the Restatement (Third) of Torts’ (2009) 44 Wake Forest Law Review 1309, 1322, 1325–6 and 1328.
It may well be that the reason why conduct is judged to be a breach provides a sound basis on which to determine whether a particular consequence of that breach should be judged to be within or outside the appropriate scope of liability for consequences. For example, if, as in Chappel v Hart, the central ground on which the defendant's conduct was held to be a breach of the duty of care was that a reasonable person would have warned about the risk of a certain consequence, this provides a straightforward ground for including such a consequence within the scope of liability even if the risk had been of extremely low probability. Conversely, the mischief to which a statutory liability is expressly targeted provides a sound reason for judging some other consequence to fall outside the appropriate scope of that liability.

But typically, outside the warning and some statutory contexts, the bearing that the nature of the breach has on the truncation (ie scope) issue is much more complex and contestable, and these qualities are masked when the judgment on the truncation issue is communicated in terms of a merely asserted 'scope of the duty'. By 2005 Lord Hoffmann had recognised the inappropriate bootstraps quality of his 'scope of the duty' approach and explicitly disapproved it, recognising that by the term 'scope of the duty' he had intended to refer to what is required to avoid breach, which is always and simply to take reasonable care in the circumstances. This question 'has nothing to do with the extent of the consequences for which the valuer is liable'.

Regrettably Australian courts often refer to the 'scope of the duty'. It is preferable for such usage to cease: where courts want to refer to the issue of breach, namely what it was that reasonable care required of the defendant in the circumstances, safer more transparent terms are available such as the 'content of the duty of care' or the 'content of the obligation'.

IV FACTUAL CAUSE

Necessary factors
Courts throughout the world are agreed that the relation of necessity (ie 'but for') between the breach and outcome is one that the law should designate as causal. If then, according to the fact-finder's evaluation of the evidence, the plaintiff has established to the requisite standard of proof that the outcome would not have occurred if the breach had not occurred; the plaintiff has shown that the breach is a factual cause of the outcome. Thus in Chappel v Hart it was because, in the evaluation of the fact-finder, it was clear that Mrs Hart would not have had the operation where and when she did had Dr Chappel given her the warning he should have given, that his breach was held

32 (1998) 195 CLR 232. See generally Stapleton, 'Occam's Razor Reveals an Orthodox Basis for Chester v Afshar', above n 18, 447-8. But even if it is completely uncontroversial that the outcome was 'the very thing' the risk of which had been a reason why the defendant's conduct was judged to be a breach, this does not relieve the plaintiff from proving factual cause: see Adeels Palace Pty Ltd v Moubarak (2009) 239 CLR 420, 442 [51].

33 Gorris v Scott (1874) 9 LR Exch 125.


a factual cause of her perforated oesophagus.\textsuperscript{36} Similarly, in \textit{March v E & M H Stramare Pty Ltd} the breach was clearly a factual cause: had the defendant not carelessly parked his truck in the mid-line of the street, the intoxicated plaintiff would not have hit it.\textsuperscript{37} The controversial issue in both \textit{Chappel} and \textit{March} concerned the normative truncation question: should this outcome of the breach be regarded as within the appropriate scope of the defendant's liability? This is not a question of fact or 'common sense' but of normative judgment on which reasonable minds might differ.

There can be more than one necessary factor, as illustrated by the facts of \textit{March} itself where both the plaintiff's intoxicated driving and the defendant's careless parking of the truck on the mid-line of the road were but-for factors of the collision.\textsuperscript{38}

\textbf{Multiple sufficient factors}

Though the but-for test of factual causation is notoriously inadequate, the Ipp Report made no recommendation on the issue of which relations besides this one of necessity should be designated as causal by the law. The grounds given for this omission were unconvincing. First it was said that the problem lay in cases of over-determination by multiple sufficient factors. An example of this is what I call the 'double hit hunters' case' where two hunters carelessly shoot and a mountain walker is hit by both bullets each of which would have been sufficient to kill instantly. In fact the problem for the law is, as we will see, much more extensive than these special cases. The second argument for inaction, namely that 'the law has devised rules for resolving such cases in ways that are generally considered satisfactory and fair',\textsuperscript{39} is wishful thinking: judicial treatment of these cases is confused, a state of affairs that is understandable given the relative neglect of causal analysis by legal scholars.

Though a regime of tort law might confine its interest to necessary (ie but-for) factors, clearly none does this. In other words, the law is interested in the possibility of imposing liability on non-necessary factors and in order to do so must recognise them as qualifying as 'causal'. The most often-cited examples of non-necessary relations that the law chooses to designate as causal are multiple sufficient factors that are independent of one another. In such cases each tortious factor is designated as causal even though not necessary:\textsuperscript{40} so, in the example in the previous paragraph, each hunter is recognised as a cause. The equivalent recognition would no doubt be afforded where the multiple non-necessary factor is an omission (which would be the case if, for example, the hunters were children and the defendants were the parents who had independently and culpably failed to control their respective children). Moreover, there are plausible reasons why the same approach should be taken to the relation to the outcome of multiple sufficient omissions which are dependent — that is where, to avoid the outcome one party would have had to act on a state of affairs set up by another but where the first omitted to act and the other omitted to set up: for example

\textsuperscript{36} '[T]he task of Dr Chappel was to demonstrate some good reason for denying to Mrs Hart recovery in respect of injuries which she would not have sustained at his hands but for his failure adequately to advise her': \textit{Chappel v Hart} (1998) 195 CLR 232, 257 [69] (Gummow J).


\textsuperscript{38} An illustration of multiple but-for factors that are omissions was given in \textit{Bennett v Minister of Community Welfare} (1993) 176 CLR 408, 429 (McHugh J).

\textsuperscript{39} Commonwealth of Australia, above n 22, 109.
where one doctor carelessly failed to place in a patient’s medical record a note of critical information concerning a patient’s medical history and a later treating doctor carelessly failed to seek the patient’s medical records.\footnote{Elayoubi v Zipser [2008] NSWCA 335. But the point has divided academic opinion: see Stapleton, ‘Choosing’, above n15, 477–9.}

\textbf{Non-necessary non-sufficient factors}

But the law is also concerned with factors that contribute to an outcome but which are neither necessary nor sufficient for it to exist. Consider the following scenarios:

- Five members of a club’s governing committee unanimously, but independently and in breach of duty, vote in favour of a motion to expel Member X from the club, where a majority of only three was needed under the club’s rules. The vote of committee member number one is neither necessary nor sufficient for the motion to pass. This is true of the vote of each member, yet the motion passed. Where there is a liability rule requiring proof that the vote of the individual voter was a factual ‘cause’ of a motion passing, the law must recognise this relation of one vote to the passage of the motion as ‘causal’.\footnote{Stapleton, ‘Choosing’, above n15, 443. The concept of a causal contribution must be carefully distinguished from the notion of ‘damage’. Suppose three votes had not been in breach of a duty: then a defendant’s vote in a breach of duty would be a factual cause of the expulsion; but would not have caused ‘damage’ to X (ie to the prospects to which X was legally entitled) because had X suffered no breaches of duty he would still have been expelled.}

- Able, Baker, and Charlie, acting independently but simultaneously, each negligently lean on Paul’s car, which is parked at a scenic lookout at the edge of a mountain. Their combined force results in the car rolling over the edge of a diminutive curbstone and plummeting down the mountain to its destruction. The force exerted by the push of any one actor would have been insufficient to propel Paul’s car past the curbstone, but the combined force of any two of them is sufficient.\footnote{American Law Institute, Restatement (Third) of Torts: Liability for Physical and Emotional Harm (2010) vol 1, 380–1 (s 27, Illustration 3) which designates the relation of each actor’s negligence to the car’s destruction as being a ‘factual cause’.}

There are many other instances where the contribution of the defendant’s breach of duty to the outcome in issue is neither necessary nor sufficient. Moreover, there may be situations where the extent of a factor’s contribution to an outcome is hard for the honest plaintiff to quantify so that, while he can show it contributed, he is unable to show that this contribution was necessary or sufficient for the outcome. This is typically the situation in pollution cases and in cases where the plaintiff has made a decision after taking into account a number of considerations. For example, consider the following:

\footnote{Compare Elayoubi v Zipser [2008] NSWCA 335. But the point has divided academic opinion: see Stapleton, ‘Choosing’, above n15, 477–9.}
Three factories each independently and in breach of duty discharge oil into a bay. By a regulatory standard, fishing in the bay is forbidden if the concentration of oil is greater than a particular level. By the time the pollution is detected the concentration level far exceeds this regulatory standard. When the ban is triggered this causes grave economic injury to local commercial fishermen who are unable to quantify the contribution each factory's discharge made.

Xavier, Yadra and Zach independently but fraudulently advise Penny that the published financial statements of a company, Tenron, are honest and strong. This advice is sufficient to persuade Penny to invest in Tenron, whose share price almost immediately collapses when it is revealed that the financial statements of the company dishonestly misrepresented the health of its finances.

What these four examples show is that a non-necessary non-sufficient factor may contribute to the existence of a phenomenon. It does so by forming part of an undifferentiated whole that operates to bring about the existence of the phenomenon: a single vote formed part of the unanimous resolution; Able's push formed part of the total physical force that sent the car off the edge of the mountain; the discharge by one factory formed part of the total concentration that exceeded the regulatory standard; and Xavier's advice was part of the total information that persuaded Penny. The examples also illustrate how the law is interested in the possibility of imposing liability on such a non-necessary non-sufficient factor. To do so the law needs to designate a notion of factual causation that is wide enough to accommodate these contributions. This is why the 'but-for' test of factual cause is under-inclusive and why courts grasp at vague undefined labels such as 'substantial factor' and 'material contribution' in their attempt to recognise a non-necessary non-sufficient factor as a 'cause'.

It is only the seductive simplicity of the 'but-for' test that distracts courts from enunciating an appropriately wide statement of the relation of 'factual cause': namely, that a factor is a factual cause if it contributes in any way to the existence of the phenomenon in issue. Courts should no longer allow the fact that in most cases the contribution in issue will have been one of necessity (but-for), to mislead them into regarding necessity as the fundamental form of causal relation recognised by the law: courts should grasp that it is the relation of contribution.

An appreciation that contribution is the fundamental form of causal relation is especially important in the context of decision-making where, as Justice McHugh emphasised in Henville v Walker:44

... the long-standing recognition of the possibility that two or more causes may jointly influence a person to undertake a course of conduct ... a representation need not be the sole inducement in sustaining the loss. If 'it plays some part even if only a minor part', in contributing to the course of action taken ... a causal connection will exist.45

Two associated points need to be mentioned here. First, notice that the four examples given so far involve an indivisible outcome and the discussion has been about a non-necessary non-sufficient contribution to the existence of that indivisible

45 Ibid 493 [107]. See also Medlin v State Government Insurance Commission (1995) 182 CLR 1 where the plaintiff's injuries from the tort were a 'contributing cause of his decision to retire' (at 8).
outcome: if the breach did make such a contribution and if all other elements of the cause of action are established, the defendants are jointly and severally liable for the entire indivisible outcome. This should be sharply distinguished from cases where the issue is whether a factor is a factual cause of only part of a divisible outcome, say pneumoconiosis or deafness: here if all elements of the cause of action are established, the defendant is only liable for that part of the divisible outcome in relation to which the breach was shown to be a factual cause (unless a special proof rule is available: see below).

Secondly, though it is for normative reasons that the law requires its notion of 'cause' to be wider than the relation of necessity (ie but-for factors) and to include any relationship of contribution to the existence of the phenomenon, whether this relationship is present in any individual case is a matter of objective fact not normative choice. Two amplifications of this latter point should, however, be noted when we move to the quite separate issue of how this factual causation relationship of contribution might be proved: the evaluative nature of assessing evidence; and the possibility of special proof rules.

The relevant evidence on causal contribution is often disputed even in the most banal of orthodox cases so that the decision-maker must make an evaluative decision on whether this factual relation of causal contribution is to be treated as having been established to the requisite standard of proof, but the evaluative nature of this process does not alter the factual nature of the underlying relation in dispute.46

The other useful amplification here is that, as we will see below, in a very few distinct areas the law constructs special localised proof rules: while these are prompted by particular normative concerns, they also do not alter the factual nature of the underlying relation in issue.

The Ipp Report
We are now in a position to consider the following Ipp Report recommendations with respect to factual causation:

(c) The basic test of 'factual causation' (the 'but for' test) is whether the negligence was a necessary condition of the harm.

(d) In appropriate cases, proof that the negligence materially contributed to the harm or the risk of the harm may be treated as sufficient to establish factual causation even though the but for test is not satisfied.

(e) Although it is relevant to proof of factual causation, the issue of whether the case is an appropriate one for the purposes of (d) is normative.

(f) For the purposes of deciding whether the case is an appropriate one (as required in (d)), amongst the factors that it is relevant to consider are:

(i) whether (and why) responsibility for the harm should be imposed on the negligent party, and

(ii) whether (and why) the harm should be left to lie where it fell.47

46 Contrast Roads and Traffic Authority v Royal (2008) 245 ALR 653, 675 [84] (Kirby J) that 'the determination of causation-in-fact is not one that can be made without recourse to broader considerations'.

47 Commonwealth of Australia, above n 22, 118 (Recommendation 29).
One lamentable effect of these recommendations is to entrench the status of the inadequate but-for test at the expense of recognising the indisputable factual involvement of non-necessary contributions (such as the individual hunter in the double-hit hunters’ case or the individual vote in the club case) in the existence of the relevant phenomenon. Thus we will have a dead mountain walker and an expelled club member but the causal contribution of, respectively, any individual hunter and any voter will not be unequivocally and coherently acknowledged, as it needs to be, because they do not fall within para (c). Instead courts are advised that non-necessary contributions to harm may be treated as a cause (para (d)) but that this should be seen as a normative choice in the individual case (para (e)). Once again factual issues have been falsely characterised as normative issues.

V PROOF OF FACTUAL CAUSE

The formulation of breach fixes what needs to be proved in the factual cause inquiry. For example, suppose a person is driving at 40 mph, which is 10 mph above what is reasonable in the circumstances, when a child darts out and, being unable to stop in time, the driver collides with the child. What should the subject matter of the factual cause inquiry be? Should it be the tortious aspect of the conduct, which is the increment of speed above what was reasonable? Or should it simply be the actor’s conduct, namely, his driving? Clearly, it is the former. The law is not interested in whether ‘driving’ contributed to the collision; it is interested in whether the breach did so.

It is sometimes forgotten that it is the formulation of breach that sets up what needs to be proved in the factual cause step of legal analysis. For example, in Roads and Traffic Authority v Royal\(^{48}\) the alleged breach by the defendant, a road authority, was that it had allowed an intersection design in which one moving car could obscure or mask the presence of another moving car from the view of a car stationary at the intersection.\(^{49}\) Yet, as the High Court (Gummow, Hayne and Heydon JJ) found, in this particular case the stationary driver had been able to see the relevant vehicle\(^{50}\) so the breach as formulated could not have contributed to the accident at all.\(^{51}\)

Next: the orthodox rule is that on the issue of factual causation the plaintiff bears the burden of proof on the balance of probabilities. The Ipp Report confirmed the wisdom of this and disapproved the recognition of special proof rules that shift the burden of persuasion to the defendant.\(^{52}\)

Sometimes this proof involves a number of steps. For example in a failure-to-warn claim against a pharmaceutical manufacturer the plaintiff typically has to establish that the drug is capable of contributing to the relevant outcome (ie that it has what we

---

49 Ibid 660-1 [26].
50 Ibid 659 [18].
51 See also Romeo v Conservation Commission of the Northern Territory (1998) 192 CLR 431, 452 [45] per Toohey and Gummow JJ:
   if the appellant established a breach of the duty of care cast upon the respondent, by reason of the failure to provide a fence a finding of causation was almost inevitable.
   If negligence lay in the failure to provide a warning sign, causation would remain a live issue.
52 Commonwealth of Australia, above n 22, 111 (paragraph 7.34).
might call 'generic capacity'\(^{53}\) and that it did so in this individual case (which we might call 'individual agency') before the critical factual cause question can be asked about whether the failure-to-warn contributed to the outcome.

Another complexity arises here. It is often easy to grasp the notion of how a contribution that is neither necessary nor sufficient may contribute to the existence of a phenomenon by forming part of an undifferentiated whole that operates to bring about the existence of the phenomenon: for example, how a single vote formed part of the unanimous resolution and how Able's push formed part of the total physical force that sent the car off the edge of the mountain (see above). But it may sometimes be helpful for courts or commentators to be able to refer to an algorithm that represents the notion of contribution. This is known as the NESS test:\(^{54}\) a factor, such as a single vote in our club example, 'contributes' to the outcome if it is a Necessary Element for the Sufficiency of a Subset of the facts. This algorithm identifies the contribution of the vote of committee member number one to the passage of the motion to expel: this vote is necessary for the subset (consisting of the vote of number one along with the votes of number two and number three) to be sufficient for passage.

Even in the typical case where the plaintiff's claim of factual cause rests on the claim that the breach was necessary for the outcome, proof of factual causation clearly involves establishing both facts about the actual past and speculation about what would have happened in the past in some hypothetical world where (inter alia) the breach had not occurred.\(^{55}\) Orthodoxy requires both to be established on the balance of probabilities.\(^{56}\) (These issues are often hotly contested requiring the decision-maker to make careful evaluations of complex evidence.) So, for example, where hypothetical conduct is in issue the relation of interest to the law is what the particular individual would probably have done. The Ipp Report correctly emphasised that the relation of interest is what the relevant individual would have done, not what some normative creation, such as a 'reasonable person', would have done: thus, in Commissioner of Main Roads v Jones\(^{57}\) the issue was what this driver would probably have done had there been relevant warning and speed signs, not what a reasonable driver would have done. In their post-Ipp Report civil liability reform legislation\(^{58}\) five States explicitly reiterated this rule in relation to the hypothetical conduct of the injured person.

But the Ipp Report and the responding legislation are significantly incomplete: the appropriate subjectivity rule is a general one and applies to the hypothetical conduct, not only of the injured person, but also of the defendant and third parties. Fortunately the High Court has been alive to the need to supplement the post-Ipp Report civil liability legislation with common law principles. Thus in Adeels Palace Pty Ltd v

---

\(^{53}\) See, eg, Seltsam v McGuiness (2000) 49 NSWLR 262 (is inhalation of asbestos capable of contributing to the contraction of renal cell carcinoma?).

\(^{54}\) See Stapleton, 'Choosing', above n\(^{15}\), 444, 459 and 471ff.


\(^{56}\) These issues must be distinguished from the contingency that another factor which did not operate might have done so and caused the phenomenon in issue (such as an injury): see Malec v JC Hutton Pty Ltd (1990) 169 CLR 638.

\(^{57}\) (2005) 215 ALR 418.

\(^{58}\) Civil Liability Act 2003 (Qld) s 11(3)(a); Civil Liability Act 2002 (NSW) s 5D(3)(a); Wrongs Act 1958 (Vic) s 51(3); Civil Liability Act 2002 (Tas) s 13(3)(a); Civil Liability Act 2002 (WA) s 5C(3)(a).
Moubarak when two patrons who had been shot by another patron sued a restaurant for providing inadequate security, the High Court correctly inquired into whether the evidence of the specific assailant, ‘a determined person armed with a gun and irrationally bent on revenge’, supported the plaintiffs’ claim that additional security staff would have succeeded in deterring him.

VI SPECIAL RULES CONCERNING PROOF OF FACTUAL CAUSATION

Plaintiffs can face insuperable evidentiary gaps when trying to establish either the relevant past facts or what would have happened in a hypothetical past situation. Of course usually such a plaintiff fails but on occasion courts or legislatures have, in response to a particular policy concern, crafted a special rule allowing the plaintiff to leap the evidentiary gap and establish factual cause.

Before analysing these, however, it is worth noting that in rare cases a special rule of proof of factual cause has been created which makes the plaintiff’s task more difficult.

Self-serving testimony

In some contexts the concern that a plaintiff will give self-serving testimony as to what they would have done in the relevant hypothetical past situation has triggered a rule barring such testimony. The Ipp Report recommended such a rule and four States have included it in their reform legislation.

What these moves overlook is that self-serving testimony may be proffered by any party. Consider a medical negligence case where the alleged breach was a failure to attend to the patient. If the plaintiff can prove that, had the defendant attended, the defendant would, on the balance of probabilities, have taken steps that would have prevented the deleterious outcome, factual cause will be established. But what if one possible reasonable treatment would not have achieved such prevention: as was the case in Bolitho v City and Hackney Health Authority. The defendant has a real incentive to give the self-serving testimony that, had she attended the patient, she would have chosen the treatment regime that would have made no difference to the outcome.

Heeding presumption

Much more commonly courts and legislatures have crafted a special rule concerning the proof of factual cause to assist the plaintiff. Sometimes such a rule relates to proof of what would have happened in the relevant hypothetical past world had there been no breach. A striking example was created in the US when plaintiffs brought defective
product claims and alleged that the defect was a failure to warn.\textsuperscript{65} When relying on such an allegation the plaintiff is required by orthodox rules to show that, had an adequate warning been given, it would have been read and heeded and the injury avoided. But how can a plaintiff prove this in the face of firm psychological evidence that suggests people rarely read labels? For policy reasons many US courts decided to assist a plaintiff over this evidentiary problem by recognising a 'heeding presumption', that the warning would have been heeded and acted upon; this is virtually always enough to get the issue to the jury.

**Market share doctrine: known mechanism; indivisible injury**

In other contexts a special pro-plaintiff rule of proof of factual causation allows the plaintiff to jump an evidentiary gap concerning past facts. An exotic example of this is the market-share doctrine under which the plaintiff need not establish that the defendant's product was the one that injured her and can recover against that defendant in proportion to its market share in relation to that product.\textsuperscript{66} In California this special rule is characterised as one merely going to the proof of factual causation and not to a reformulation of what can form actionable damage, that is, what can form the 'gist' of the claim.\textsuperscript{67}

**Alternative liability: known mechanism; indivisible injury**

The most well-known rule relating to an evidentiary gap concerning past facts is the 'alternative liability' rule from *Summers v Tice*,\textsuperscript{68} which has been adopted by virtually all US jurisdictions. Here there were two hunters: the mechanism of injury was known to be due to just one agent (only one shot had hit the victim); the injury was known to be indivisible; the number of tortfeasors was small (two); all tortfeasors were before the court; and identification of the involved agent was impossible. Each was held jointly and severally liable.

**The 'indivisibility-of-injury' rule: known mechanism; cumulative injury**

In some jurisdictions\textsuperscript{69} a different pro-plaintiff rule of proof about past facts has been created in certain cases where the total injury was known to be divisible but the relative contributions to the total injury could not be adequately determined so that while it was known that each agent had tortiously caused some but not all the total injury, the defendants were nevertheless held jointly and severally liable for the total

\textsuperscript{66} See Sindell v Abbott Labs, 607 P2d 924, 937 (Cal 1980); Brown v Superior Court, 751 P2d 470, 485-7 (Cal 1988).
\textsuperscript{67} See Jolly v Eli Lilly & Co, 751 P2d 923, 930 (Cal 1988). In New York the doctrine cannot be characterised this way because there it is no answer to the claim for the defendant to prove the relevant unit could not have been one he produced: see Hymowitz v Eli Lilly, Co, 539 NE2d 1069, 1078 (NY 1989).
\textsuperscript{68} 199 P2d 1 (Cal 1948).
injury as if it was indivisible. In the United States courts commonly recognise such an ‘indivisibility-of-injury’ fiction in relation to water pollution and asbestosis.\textsuperscript{70} While the result in Bonnington Castings Ltd v Wardlaw\textsuperscript{71} is also consistent with the ‘indivisibility-of-injury’ fiction, it is unclear whether the House of Lords intended to create such a special rule of proof of factual cause because the case revolved around a more fundamental issue, namely whether a defendant could be liable at all if his breach had only caused part of a total injury.\textsuperscript{72} When the point came for resolution in other cumulative disease cases in the UK\textsuperscript{73} no special rule was applied: the plaintiff was required to show, on a rough and ready basis, how much of the total injury the breach had factually caused.

The material contribution/exposure to risk doctrine: unknown mechanism; indivisible injury

Much more radical is the rule created in some jurisdictions to allow plaintiffs to establish that one source of risk among many was a factual cause of an indivisible medical condition when the mechanism of that condition is unknown, as is the case with mesothelioma. A victim of such a condition could not succeed on orthodox principles because it is not possible to make a ‘robust and pragmatic’\textsuperscript{74} inference that the individual source of risk contributed to the condition: an impossibility elegantly exposed in Fairchild v Glenhaven Funeral Services Ltd.\textsuperscript{75} Nor is it possible to apply the special Summers v Tice proof rule of alternative liability which requires all risk-creators to be culpable and before the court; nor can the plaintiff seek to use the ‘indivisibility-of-injury’ fiction which requires that it is known that the total injury resulted from cumulative injuries from every source of risk.

The radical rule of proof adopted in these unknown-mechanism cases, described as the ‘material contribution to risk’ doctrine in the UK or the ‘exposure to risk’ doctrine in the US, does present significant challenges, for example: what is its incidence (eg does the rule only apply to ‘single-agent’ conditions\textsuperscript{76})? Does the rule shift the burden of persuasion? Does the rule support in solidum or joint and several liability? And is the rule appropriately characterised only as a special rule of proof of factual causation or is it preferable to see it as recognising exposure to risk (ie loss of a chance) as actionable

\textsuperscript{70} See Stapleton, ‘The Two Explosive Proof-of-Causation Doctrines’, above n\textsubscript{6}. Both the ‘indivisibility-of-injury’ rule and the ‘alternative liability’ rule have been restated in s 28 of the Restatement (Third) of Torts: Liability for Physical and Emotional Harm (2010). The former had been covered by s 433B(2) of the Second Restatement while the latter had been covered by s 433B(3).

\textsuperscript{71} [1956] AC 613.


\textsuperscript{73} Such as Thompson v Smiths Shiprepairers (North Shields) Ltd [1984] 1 QB 405 (deafness) and Holby v Brigham & Cowan (Hull) Ltd (2000) 3 All ER 421, 428 (asbestosis).

\textsuperscript{74} Wilscher v Essex Area Health Authority [1988] AC 1074, 1090 per Lord Bridge in a doomed attempt to fit McGhee v National Coal Board [1973] 1 WLR 1 into orthodox principles.

\textsuperscript{75} [2003] 1 AC 32. See, eg, Lord Bingham at 57–8 [22].

damage in the limited area of incidence of the rule? Nevertheless, turning a blind eye to the dilemma that such diseases present to orthodoxy leads to the sort of incoherence that bedevils the case law on asbestos cancers in Australia which currently proceeds on the basis of unacknowledged and conflicting fictions about aetiology.

The Ipp Report

We are now in a position to consider the Ipp Report recommendation in relation to proof of factual causation that 'in appropriate cases, proof that the negligence materially contributed to the harm or the risk of the harm may be treated as sufficient to establish factual causation' so long as the normative nature of such a special rule of proof is acknowledged and openly analysed. The recommendation is to be welcomed in as much as it will bolster the separation of the factual question of whether a factor is a cause from the normative issue of whether this particular class of plaintiff should be afforded a special rule of proof on this factual issue.

Five States have enacted a provision broadly in line with this recommendation but South Australia and the ACT limited their provision to cases where the plaintiff 'has been negligently exposed to a similar risk of harm by a number of different people (the defendants) and it is not possible to assign responsibility for causing the harm to any one or more of them'. The latter provision is clearly not wide enough to authorise special rules in one-wrongdoer cases such as a heeding presumption or the approach adopted in cases such as McGhee v National Coal Board. This limited provision may not even authorise a rule purporting to cover situations where not all risk creators need be joined as defendants such as the market share doctrine, the 'indivisibility-of-injury' rule and the material contribution/exposure to risk doctrine.

But such developments may still occur at common law. Apart from the entrenchment of the rule that the burden of proof rests on the plaintiff, there is nothing in the post-Ipp Report civil liability legislative provisions dealing with either factual causation or scope of liability that inhibits the future recognition of special rules of proof or the enunciation of guiding principles governing the scope of liability. In relation to both, the legislation is best seen as reporting the legitimate law-making role of the courts, save with the caveat that courts must enunciate the considerations that have been taken into account. Beyond that the legislation properly 'offers no further guidance about how the task is to be performed.' In these areas, and despite

77 Note the problematic characterisation by Lord Hoffmann in Barker v Corus UK Ltd [2006] 2 AC 572 on which see Stapleton, 'Occam's Razor Reveals an Orthodox Basis for Chester v Afshar', above n 18.
79 Commonwealth of Australia, above n 22, 118 (Recommendation 29(d)).
80 Civil Law (Wrongs) Act 2002 (ACT) s 45(2); Civil Liability Act 1936 (SA) s 34(2).
82 Jane Stapleton, 'The Golden Thread at the Heart of Tort Law: Protection of the Vulnerable' in Peter Cane (ed), Centenary Essays for the High Court of Australia (2004) 242, 244: A vital secret of our constitutional arrangements is the close union of the judicial and legislative powers in the court of ultimate appeal and that our common law legal systems embrace a form of the separation of powers doctrine that accommodates this substantial law-making capacity.
83 Adeels Palace Pty Ltd v Moubarak (2009) 239 CLR 420, 443 [54].
superficial variations in statutory language, the critical role of the High Court in nurturing and developing a national system of common law remains undiminished.

THE FUTURE

Australian courts have struggled to express the distinction between factual causation and the truncation (ie the scope) issue, a struggle compounded by the lamentable recourse to the slogan of 'common sense causation'. The attempt by the Ipp Report and the responding legislation to achieve the separation of these two issues was undermined by the retention of the umbrella term of 'causation' to cover the amalgam of both. Since that umbrella term does no substantive work in the legislation, Australian courts should quietly ignore it. They should also no longer obscure their judicial reasoning by reference to 'common sense causation'. A great advantage of such moves is that it encourages exposure of the nature, variety and complexity of concerns in play at the truncation stage.

If then, the obfuscation of 'common sense causation' is stripped out of future truncation analysis what might we find? It is obvious that, just as with the issue of what 'reasonableness' requires on the facts of a case, the scope issue cannot be reduced to some formula. On the other hand, just as with 'reasonableness', the scope issue does contain some internal structure which we can begin to enunciate. For example, we can say that a consequence will fall outside the appropriate scope of liability for negligence unless it at least: can plausibly be said to fall within the 'perimeter rule' of 'foreseeability of the type of harm';84 'damage' relative to the normal expectancies of the plaintiff absent torts;85 is not a coincidental consequence;86 and is the result of one of the risks that made the conduct careless.87 Moreover, in judging where the chain of responsibility should be truncated a court may take account of a range of other factors such as concern with disproportion88 and attenuation, or a concern to shield a particular class of defendant.89 In other words, the biggest payoff of separating factual cause from scope and abandoning the slogan of 'common sense causation' is that we can get to work on understanding what principles, policies and concerns govern the scope issue.

85 Stapleton, 'Cause-in-Fact', above n 4, 401, 412–17. Travel Compensation Fund v Tambree (2005) 224 CLR 627 runs counter to the usual judgment that if, but for the breach of an obligation of care, the plaintiff would have suffered an equivalent loss in a different transaction, it lies outside the appropriate scope of liability.
86 Stapleton, 'Ocam's Razor Reveals an Orthodox Basis for Chester v Afshar', above n 18, 438ff.
88 See, eg, Homac Co v Sun Oil Co, 180 NE 172 (NY 1932).
89 Stapleton, 'Cause-in-Fact', above n 4, 420–1.