A CRITICAL LOOK AT DETERMINATE THEORIES OF CAUSATION IN LAW

by

ERIC ADAMS

B.A., The University of Calgary, 2003
L.L.B., The University of Victoria, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

MASTER OF LAWS

in

THE FACULTY OF GRADUATE STUDIES

(Law)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2009

© Eric Adams 2009
Abstract

My thesis critiques three determinate theories of causation in law: (1) the neutrality of Richard Wright's test for "causes-in-fact"; (2) the use of "abnormality" to distinguish causes from mere conditions in Hart and Honoré's ("H&H") theory of proximate causation; and, (3) the substitution of "proximate causation" with an efficiency calculus in the Law and Economic literature. The first chapter argues that the way a causation question is formulated will largely determine the "facts" we receive from "necessity" or "sufficiency" tests for causation-in-fact. In the second chapter of my thesis I explore the concept of "normality" in H&H’s theory of proximate causation in Causation in the Law. I explore how H&H’s "normality" implies "normalization", serves as a standard of conduct and affects substantive outcomes. The third chapter offers a case study for the second. It argues that Canadian courts tacitly weigh risks against rewards when constructing the meaning of the term "accident" in coverage decisions. This demonstrates how courts make policy choices when ostensibly applying an "empirical" standard of proximate cause. The fourth chapter considers the substitution of efficiency for proximate causation in the Law and Economics literature. It argues that efficiency cannot be reduced to a matter of empirical fact.
Table of Contents

Abstract........................................................................................................................................ii

Table of Contents .......................................................................................................................... iii

Acknowledgements ......................................................................................................................... v

Introduction...................................................................................................................................... 1
   A Note about Organization ........................................................................................................... 9

Wright’s NESS Test..........................................................................................................................12
   NESS Causation.............................................................................................................................12
   Substantial Factors and Material Contribution ...........................................................................15

The Overstated Advance of NESS Causation ..............................................................................17
   Preemptive Causation....................................................................................................................17
   General and Specific Descriptions ...............................................................................................20

Indivisible Entities and Uncertain Causal Processes ....................................................................24
   Indivisible Consequences.............................................................................................................25
   Indivisible Causes..........................................................................................................................27
   Asymmetrical Overdetermination ..................................................................................................28
   Thresholds....................................................................................................................................32
   Omissions......................................................................................................................................34

Conclusion........................................................................................................................................35

Hart and Honoré’s Normality Standard .........................................................................................38
   Normality as Custom.....................................................................................................................41
   Normality and Sine Qua Non .........................................................................................................42
   Normality and Normalization ........................................................................................................48

Normality as Ordinary Consequences .......................................................................................52
   Original and Intervening Causes ..................................................................................................52
   Normality, Forseeability and Natural and Probable Results .......................................................54
   Normalizing Contingencies ...........................................................................................................56

Conclusion........................................................................................................................................58

Finding the “Normative” in the “Empirical”: Examples from Insurance Law .........................60
   The Definition of “Accident” in Canadian Coverage Cases and the Unspoken “Useful Purpose” Test.........................................................................................................................61

The Expectations Approach...........................................................................................................63
   The Subjective Standard.................................................................................................................63
   Objective Standard Cases .............................................................................................................65
# The “Useful Purpose” Test

Purpose and Credibility

“Expectations”

Examples from the Case Law

Conclusion

## Instrumental Causation, Efficiency and Insurance Law

**Variety and the Law and Economics of Liability Insurance**

Why Insurance?

**Variety and Rational Risk-Taking**

Variety and Tort Standards

Variety and Insuring the ‘Monstrous’

Contingent and Hedonic Valuation

Wealth Maximization

**Variety and Irrational Risk-Taking**

Bounded Rationality

Variegated Rationality

Prescribing Irrationality

Paternalism

Conclusion

**Bibliography**
Acknowledgements

I would like to express my deep gratitude to my supervisor, Bruce MacDougall, for his invaluable encouragement, direction and supervision throughout this research work. His support and thoughtful advice allowed me to complete this project successfully. I am also indebted to Cristie Ford, Anthony Saunders and Mary Liston for their suggestions and kind support.
Introduction

When we say in law that “X caused Y” is this statement made strictly on an assessment of objective facts or does the statement also involve questions of policy? My thesis argues the latter. Academic writing on legal causation might be loosely organized under one of three camps. The first maintains that finding causation in law involves the application of universal causal rules which determine the issue in any particular case. Their writing often employs the language of “chains of causation”, “intervening events” or other causal heuristics that traditionally inhabited common law causation decisions until their slow retreat in the first half of the twentieth century. The strongest articulation of this thesis would analogize the process of determining legal causation to the way that causes are discovered in the natural sciences. A weaker version allows for a distinction between the study of causation in law and in science, but maintains that legal causation is not contingent or dependent upon public policy or instrumental concerns. Of this latter group, Hart and Honoré's invocation of common sense as a guide to determining causation represents the best known and most influential example.


My position is largely situated at the other end of the spectrum. At this end, the scholarship maintains that the selection of a cause in law is policy-laden and dictated by practical or instrumental concerns. This group includes Legal Realists who argue that the rhetoric of causation decisions operates to mask judicial preferences and CLS successors who stress the political linkages between objective causal theories and liberal decision-making. It is also, notably, represented by a Law and Economics scholarship insisting on the instrumental, wealth-maximizing functions of causal determinations.

The third camp sits between the other two. They distinguish the determination of (possibly numerous) “scientific causes” or “causes-in-fact”, from the determination of the “proximate cause” or “cause-in-law” responsible and liable for a legal injury. The former question is answered using objective and universal principles of causation, while the latter is determined on policy or instrumental considerations.

My thesis targets three arguments within this spectrum: (1) the neutrality of Richard Wright’s

---


test for “causes-in-fact”; (2) the use of “abnormality” to distinguish causes from mere conditions in Hart and Honoré’s (“H&H”) *Causation in the Law*; and, (3) the substitution of “proximate causation” with an efficiency calculus in the Law and Economic literature. My thesis is meant to question the determinate nature of these causal analyses and to highlight instances of normative concern. The chapters proceed from fact to policy—from considering theories of objective causation to instrumental causation. The doctrine of causation-in-fact, as the name suggests, is most widely, though not unanimously, accepted as being grounded in objective fact.9 The literature is largely divided, on the other hand, on whether the selection of a “proximate” or “responsible” cause proceeds from factual or policy-based considerations.10 As noted above, this was a major point of contention for the American Realist movement and the site of one of its most lasting successes. In Law and Economics the question of causation is secondary or subordinate to policy (efficiency) concerns. These three arguments considered in this thesis are selected for their importance to determinate theories of causation-in-fact, proximate causation, and instrumental causation (efficiency theory), respectively.

Wright's contribution to the literature is important, if not as much for braving new causal territory, for significantly clarifying causation issues and theoretical positions. His contribution is twofold. First, Wright found and isolated (and coined the term) the NESS test from the rich discussion of common sense and linguistic causation in H&H's *Causation in

---


10 Or, between “causal maximalists” and “causal minimalists” in H&H's jargon, *supra* note 4 at lxxiv.
the Law. The test is as follows:

an element is considered a contributing cause of an event if it represents a necessary element of a set of conditions actually operating at the time of a loss that are sufficient though not necessarily necessary to produce the loss.11

Wright declares this test as an advance on the “but for” or sine qua non test12 for its ability to handle situations of causal overdetermination. Causal overdetermination occurs in situations where more than one set of factors involved in a loss would each have been independently capable of causing the loss and, thus, were not, in themselves, necessary for the event. An example is where two fires enjoin to consume a house notwithstanding that both would have been capable of destroying the house in the absence of the other.13 Wright’s second contribution is his stalwart insistence on the distinction between the NESS analysis and the doctrinal or ad hoc policy considerations which narrow the list of NESS factors down to those deemed liable for the loss. While Wright believes the narrowing process requires policy choices, he maintains that the selection of “causes-in-fact” is an empirical matter that is both objective and apolitical.

The first chapter of my thesis questions Wright’s strict division between an objective NESS “causation-in-fact” analysis and a policy-laden narrowing process. I argue that there are essentially two situations where the law might viably substitute a necessity for a sufficiency

11 Wright, “Causation in Tort Law”, supra note 8 at 1788-1813; Wright, Pruning, supra note 8 at 1019.
12 The terms, “but for” and sine qua non are used synonymously and interchangeably throughout my thesis.
13 Anderson v. Minneapolis Ry. (1920) 146 Minn. 430. Discussed, for example, by H&H, “Causation in the Law,” supra note 4 at 202; Wright, “Causation in Law,” 1792-93 supra note 8; Malone, Ruminations, supra note 5 at 89.
(a) where the law must attribute responsibility for a legally indivisible entity that, in nature, is composed of infinitely smaller organic constituent parts (such as a house destroyed by fire); or
(b) where it is impossible for a court to determine the necessary causes in a particular process because the process is too complex or microscopic for explication in a courtroom.

The utility of sufficiency or necessity as “causation-in-fact” heuristics depends largely on how a causal question is framed. The framing dictates the tools we use, and ultimately affects the answers we receive. Accordingly, I argue there can be no strict division between the process of illuminating the causal facts and the facts themselves.

In the second chapter of my thesis I explore the concept of “normality” in H&H's *Causation in the Law*. H&H suggest that the *sine qua non* “causes” of an event are narrowed down to responsible causes by criteria rooted in our shared linguistic and common sense understandings of the relationship between causation and responsibility. A “cause” is understood as that which disturbs the *normal* course of things.14 Where a course of conduct upsets the regular cadence of social life any attending losses are attributed to the responsible agent. I argue the concept of “normality” is employed in the text in two ways. In their introductory chapters, H&H suggest that “causes” are distinguishable from “mere conditions” by their “abnormality”— or, their conspicuousness when considered against the “backdrop” of the routines or affairs of ordinary life. In a later chapter on the application of their theory

14 H&H, “*Causation in the Law*”, *supra* note 4 at 33-40.
to the law of tort, however, “abnormality” represents an unforeseen “intervening event” which severs the causal relationship between a voluntary action and its abnormal consequences.

I argue that these two uses of the concept of “normality” approximate two different ways of measuring conduct or consequences. “Normal”, in the second use of the word, might be applied objectively or without becoming involved in normative considerations or arguments. The chain of events can be assessed as an empirical matter. If a chain proceeds normally, it is attributable to the original actor. If it does not, it is attributable to the intervening actor. “Normality”, in this sense, does not necessarily imply “normalization”. In the first sense of the term, however, human conduct is measured and benchmarked against social (or even personal) norms. If such norms are embedded in legal standards, the judiciary becomes involved in their preservation or promotion. Where human norms serve as a standard of reference in law they perform a governing role. An individual is impelled to conduct him- or herself with reference to this standard. This is of normative concern. If the determination of causation implies governance through a “normality” concept, it is very difficult to maintain the separation between objective causation and policy.

The third chapter of my thesis offers something of a case study for the second. This chapter considers how courts determine whether a loss was caused by “accident” for the purposes of construing insurance coverage. The legal test for “accident” is whether an insured “expected” the loss. If he or she did not, the loss is deemed “accidental”. The issue is analogous to the determination of “proximate cause”. Where the court is asked whether a
loss was “caused” by accident, the *sine qua non* status of the insured is generally not in
doubt. The question before the court is whether the correlation of events which included the
insured's conduct is best labelled an “accident” or, rather, is attributable to the insured. This
process looks something like it does in H&H's *Causation in the Law*. The law must offer an
attributive explanation of the events by considering the nature of the *sine qua non*
contributions and how they interrelate. Insurance law does not use the exact same limiting
device as H&H. Canadian courts determine whether something was “caused” by “accident”
on the basis of an insured's *expectations* instead of a sequence's *normality*. On the other
hand, both limiting devices are ostensibly empirical rather than normative.

The third chapter argues that Canadian courts tacitly weigh risks against rewards when
constructing the meaning of the term “accident”. They inject the factual assessment of an
insured's expectations with concern as to whether the type of risk-taking in question ought to
be insurable. As suggested, this occurs tacitly. This chapter does not consider how norms
*normalize*, as is suggested in chapter two. It does, on the other hand, demonstrate how courts
decide policy when ostensibly applying an "empirical" causation standard. The presence of
this sort of dynamic represents a challenge for “causal maximalists” like H&H who insist that
the determination of “proximate cause” is not something governed by or subordinate to
instrumental concerns.

The weighing of expected losses and benefits is closely associated with the Law and
Economics literature. If my analysis in chapter three is correct and courts do weight the costs
and benefits of risk-taking in coverage decisions, this would seem to support or vindicate an
economic analysis. A cost-benefit analysis is not causal. An event can be no more said to have causally contributed to an outcome if it has redeeming economic value than if it doesn't.

On the other hand, the Law and Economics literature, as discussed in chapter four, is comfortable subordinating the issue of causation to an efficiency calculus. This calculus might inform the determination of the issue of causation or simply substitute for causation. Either way, the notion that causation doctrine operates to anchor attributive judgements to the mechanics of the world of facts is discarded in favour of the unabashedly instrumental promotion of value-creation.

This efficiency calculus, of course, first requires normatively justifiable metrics of value. If the first two chapters question theory which reduces causation to a matter of empirical fact, the fourth chapter questions whether economic value can, in place of “causation”, be similarly reduced to a matter of empirical fact. It also considers how the evident variety in the way in which people manage risk bears on the question of how we construct and weigh value.

It outlines an orthodox economic analysis of the construction of the term “accident” and proposes three hypothetical models for reconciling differences in risk-taking with rational choice theory. It discusses the problems we face if variegated models of rationality are used to inform legal standards. It argues that the law generally does not afford a credit to people deriving a special, irregular utility from risk-taking and that this result is as much due to principles of equity and fairness as efficiency. Second, the chapter discusses the normative questions which behavioural law and economics must address if persons who act outside of
social norms are deemed less-than-fully rational. It argues that the law faces a challenge in responding to differences in capacity or rationality because it is often prohibitively difficult for a court to determine at any moment whether an actor is influenced by a defensibly rational self-interest or by an irrational or boundedly-rational self-interest. It, moreover, suggests that the law often grants clemency to those suffering from cognitive biases, where behavioural economics would seem to require the opposite. This suggests that the law's concern for moral proportionally is not simply subordinate to efficiency commitments. Finally, it offers a brief comment on the need for behavioural economists to formulate normatively acceptable metrics of economic value.

A Note about Organization

The three causal theories critiqued in my thesis represent a slim selection from the very large and rich world of causation in law. The selection is limited due to the time and space constraints of this project. Wright's work is selected as a focus for my discussion on sine qua non causation because of the importance of the NESS test in the literature. I discuss H&H's work as they are the most well-known proponents of an integrated factual theory of proximate causation. The survey of the Law and Economics literature is broader. Steven Shavell's model of tort and insurance is used as a starting point for my discussion on instrumental causation due to its simplicity and accessibility. However, the discussion branches out from there. Nevertheless, Law and Economics is a very diverse literature and my discussion only touches on a fraction of its important contributions.

These theories are admittedly discussed and explored in an asymmetrical manner. The
chapters devoted to exploring Wright's NESS theory and to Law and Economics are strictly theoretical. I do not look into the cases for support. The doctrinal exploration of H&H's *Causation in the Law* and the theoretical discussion of Law and Economics are discussed in the context of insurance law.

Had I had more time, I may have found it fruitful to explore examples of NESS causation in insurance law. On the other hand, insurance law represents an interesting field to discuss both H&H and Law and Economics for particular reasons. For one, as noted above, the determination of whether an incident was “caused” by “accident” is largely analogous to the selection of a “proximate cause”. “But for” causation is generally clear. Judges proceed on the assumption that the determination of whether something occurred by “accident” is available on the facts. This assumption loosely, though imperfectly, reflects how causation is understood in H&H. Insurance law, thus, provides fertile ground for studying factual proximate causation theory. Insurance law is also a compelling area to study economic valuation. As will be discussed more fully below, people pay differentiated premiums in order to be able to participate in risk. I proceed, from this observation, to ask whether they do so because they value risk differently or because they are differently disposed to risk. This variety and the many ways it can be accommodated by law or in economic models puts the determinacy of Law and Economics into doubt.

I should note, finally, that the discussion of Law and Economics does not explore the case law for support. While I was tempted to do so, many legal economists suggest economic considerations operate beneath the rhetorical surface of the law. Books like Posner's
*Economic Analysis of Law*\(^{15}\) provide examples from a wide range of cases of how common law judges tacitly intuit their way to economically efficient outcomes. There is a certain challenge in tackling theory which claims to operate beneath the surface of things and then uses these things as evidence. Whether cases used to support economic arguments might also be used to put those arguments into doubt is an interesting question, but one that is not taken up here. I only note it parenthetically. At any rate, the discussion of the Law and Economics of liability insurance is strictly theoretical.

Wright's *NESS Test*

NESS Causation

An element is considered a contributing cause of an event if it represents a *necessary element of a set of conditions actually operating at the time of a loss that are sufficient though not necessarily necessary to produce the loss*.\(^{16}\)

In the seminal article “Causation in Tort Law”, Wright boldly declares that his *necessary element of a sufficient set* test “not only resolves but also clarifies and illuminates the causal issues in the problematic causation cases that have plagued tort scholars for centuries”\(^{17}\) and that it “is the essence of the concept of causation”.\(^{18}\) The suggestion that the NESS test resolves all outstanding causal issues is overstated. As this chapter argues, the test provides a way for judges to allocate responsibility in the face of prohibitively complex or indeterminate causal processes. It is, as I argue, a heuristic of exigency suited particularly to legal problem-solving rather than a panacea for all our causal dilemmas.

Since Hume, the modern approach to the question of causation has been to disregard the possibility of unobservable “forces” or “powers” which invisibly impel causal processes in favour of studying causation as an empirical matter.\(^{19}\) Rather, Hume argued, causation is only intelligible through generalizing from our empirical evidence. Where two things occur

---

\(^{16}\) Wright, “Causation in Tort Law”, *supra* note 8 at 1788-1813; Wright, *Pruning*, *supra* note 8 at 1019.

\(^{17}\) Wright, “Causation in Tort Law”, *supra* note 8 at 1802.

\(^{18}\) *Ibid* at 1805.

\(^{19}\) David Hume, *A Treatise of Human Nature* (Clarendon Press, 1888) at 167. Hume argued that in order to understand causal relationships we ought not to look to uncover something inherent to the properties of substances interacting in any particular occurrence or, in his words, the “power and necessity... in objects”.

12
in “constant conjunction” within a “regular sequence” the factors come to be experienced as “cause” and “effect”. This thesis was revised and reformulated by J.S. Mill in *System of Logic*. Mill accepted that causation is not discoverable from innate properties or forces operating within or on the elements of an event. He rejected, on the other hand, the idea that the “conjunction” between elements to a sequence need be universal and “constant” for their relationship to be considered causal. Rather, he suggested that a *plurality of factors* or sets of conditions can bring about an occurrence. A “cause” does not need to precede an “effect” in every set of circumstances. Causation can be established if the preceding element was necessary for the event in a *particular* set of circumstances. The theory developed in *System of Logic* has left us with the *sine qua non* test used to determine causation in law today.

The NESS test extends the *plurality* thesis further. While Mill argues that a plurality of factors or circumstances can produce a *type of outcome*, H&H and Wright suggest that there can be a plurality of circumstances that operate simultaneously and independently to produce a particular occurrence *in a specific instance*. There are a number of famous examples in the cases and literature:

- A fire is set by the engine of a train belonging to one railroad company. This fire

---

20 Hume argued that, because the human mind has no way to interpret this “necessity” in objects, causal proofs are felt rather than found. The “experience” of causation, he suggests, is the most we can expect to discover about “causes” and represents the epistemic limits of causal inquiries. *Ibid* at 247.


22 For example, *Prosser and Keeton on the Law of Torts* reads: “The defendant's conduct is a cause in fact of the event if the event would not have occurred but for conduct; conversely, the defendant's conduct is not a cause of the event, if the event would have occurred without it.” *Prosser and Keeton on the Law of Torts* § 41 (5th ed.) (West Group, 1984) at 266.
unites with a fire set by the engine of another railroad company and jointly destroys a plaintiff’s property.\(^{23}\)

- Two motorcyclists drive noisily past a horse-drawn wagon. The sound of the motorcycles spooks the horse and causes the plaintiff injury.\(^ {24}\)

- Two hunters go grouse hunting together knowing that there are other hunters in the area. Some grouse take flight and both hunters simultaneously take a shot. Both hunters hit a fellow hunter who instantly dies. It is impossible to say which bullet struck first.\(^ {25}\)

- The accumulation of five units of pollution is sufficient to cause an injury. Seven defendants, acting independently, each discharge one unit of pollution.\(^ {26}\)

In these cases, our intuitions would seem to suggest that the defendants were “responsible for” the respective injuries and should be held accountable. It would seem counterintuitive to suggest that neither of the fires, or neither of the bullets, for example, caused the losses. On the other hand, the traditional *sine qua non* test would relieve all these defendants of liability. We are not able to say that, “but for” the negligence of the first railroad company, the plaintiff’s property would not have succumbed to fire. If the first engine hadn’t sparked a fire, the property would have been destroyed by the fire sparked by the second. Neither of

\(^{23}\) Discussed by Malone, *supra* note 5 at 90; Wright, “Causation in Law”, *supra* note 8 at 1792. The hypothetical is based upon the facts of the case of *Anderson v. Minneapolis, St. P. & S. Ste. M. Ry.* 146 Minn. 430, 179 N.W. 45 (1920); *Kingston v. Chicago & NW Ry.* (1927) 191 Wis. 610, 211 NW 913.


\(^{25}\) Modified from the facts of *Cook v. Lewis*, [1951] S.C.R. 830 (S.C.C.) and *Summers v. Tice* (1948) 33 Cal. 2D 80. In *Cook* and *Summers*, however, the courts were unable to ascertain which of the bullets struck.

\(^{26}\) Wright, “Causation in Law”, *supra* note 8 at 1793. See, for example, *Michie v. Great Lakes Steel Division National Steel Corporation* 495 F2d 213 (6th cir. 1974)
the motorcycles was necessary to stir the horse. If the first hadn’t, the second would have and *vice versa*. If the first hunter’s bullet hadn't fatally struck the victim, the second would have and *vice versa*. Similarly, for each individual polluter. The strict application of a *sine qua non* test would leave these injured plaintiffs without recovery, even though the sources of their injuries are more or less clear.

*Substantial Factors and Material Contribution*

In response, American law often adopts a “substantial factor”\(^{27}\) test and Anglo-Canadian law a “material contribution”\(^{28}\) test. These tests supplement the shortcomings of “but for” causation. Where a “but for” test would counterintuitively exculpate a negligent defendant, a court can instead ask whether a force was a “substantial factor” or “materially contributed” to the circumstances of a loss. The test seems intuitively clear. Both fires, both gunmen and both motorcycles can easily be said to have “materially contributed” as “substantial factors” to their respective harms. On the other hand, it is far from clear what the phrases “material contribution” or “substantial factor” actually mean. Wright has attacked the *substantial factor* test as vacuous rhetoric that provides little instruction to judges in hard cases.\(^ {29}\) The phrases seem to imply that a factor is “substantial” or “materially contributes” if it helped cause an occurrence. The tests, on the other hand, are meant to advise as to what causation means in these types of circumstances. A factor which helps *cause* an occurrence, is thus

---


\(^{29}\) Wright, “Causation in Law”, *supra* note 7 at 1782-4.
considered one of its *causes*. The language is clearly circular.\(^{30}\)

Imagine that there was a third gunman who, at the time the other gunmen fired, simply shot a blank round. Imagine that at the time the two fires engulfed the plaintiff’s property the local firefighters were sleeping on shift, but that the fire station was an hour away by fire truck and that the plaintiff’s property was completely destroyed in less than an hour. Or that a third vehicle passed the wagon at the same time as the motorcycles, but this vehicle was a noiseless electric car. We would intuitively say the blank round, the sleeping firefighters and the electric car did not “materially contribute” as “substantial factors” to those losses. But why is this so? Normally we could say with confidence that they are not causes because they were not *necessary* ingredients in bringing about the result. If the third gunman hadn’t fired the black round, death still would have ensued. If the firefighters hadn’t been sleeping, they still would have arrived at the scene too late to save the building. If the noiseless vehicle hadn’t driven by, the horse still would have been spooked.

The NESS test provides a means to deal with such instances. We can say that the third gunmen, the sleeping firefighters or the electric car are not “causes” because they were not only *unnecessary* to produce the losses in the circumstances but that they were also *insufficient* to do so. The spark from the first engine, however, was a necessary element of starting the first fire and was *sufficient*, in itself, to bring about the loss. The spark was, thus, a “cause”. Similarly, the first motorcycle or the first shot were necessary parts of sufficient sets and, thus, causes on application of the NESS test.

\(^{30}\) *Ibid.*
The Overstated Advance of NESS Causation

Preemptive Causation

The principal virtue of the NESS test is its ability to make up for the shortcomings of the “but for” test where there appears to have been multiple sufficient cases. Wright also declares the NESS test as an advance for its ability to resolve instances of “preemptive causation”.

Examples in the literature include:

- Two assassins independently plot to kill a desert traveller. The first poisons the water in the traveller's canteen. The second, not knowing about the poison, empties the canteen. When the traveller opens her canteen to have a drink, she finds it empty and ultimately dies of thirst.  

- A driver picks his car up from a garage. The garage, however, did not repair his brakes as required and they remain inoperable. The driver approaches an intersection, but does not attempt to use the brake pedal. He strikes the car ahead of him.

In cases of “preemptive causation”, one set of factors that would have been sufficient to produce a harm is “preempted” by a second set of factors that does produce the harm. The attempt of the first assassin is preempted by the actions of the second. The traveller dies of

\(^{31}\) In Wright's words, the “duplicative causation” cases. Supra, note 8 at 1791


thirst, not of poison.\textsuperscript{34} Similarly, the negligence of the garage in repairing the brakes is said to have been preempted by the negligence of the driver. Notably, in both examples, the preempted set of factors was \textit{sufficient} to produce the loss. The NESS test, however, is more specific. It requires that the set of conditions sufficient to produce the loss \textit{actually be operating} at the time of the loss. If a man dies in a hospital and a nuclear attack kills everyone in his town the next day, we wouldn't say that the nuclear attack, though sufficient to kill him had he survived, was a “cause” of his death. It was not \textit{actually operating} at the time. Where this qualification is added, the NESS test can be used to resolve the circumstances above.

Wright's declaration of an advance on \textit{sine qua non} causation, here, however is unfounded. The \textit{sine qua non} and NESS tests operate in the same way in cases of “preemptive causation”. It is true that if a preempting factor didn't occur, a preempted factor would. Thus, neither factor would be \textit{necessary} for the occurrence. Suppose a man is fatally shot and lies bleeding on the street, beyond the hope of treatment. As he lay dying another attacker stabs him until he is dead.\textsuperscript{35} If we ask who is responsible for the victim's \textit{death} the second attacker seems to be exculpated. The man would have died due to the shooting regardless of the actions of the second attacker. On the other hand, if we ask who was responsible for the \textit{death in question} or his \textit{death under the circumstances} it is clear that both assailants are responsible. The victim would have survived longer and not met the same

---

\textsuperscript{34} Though, see dissenting opinion of Stapleton, \textit{supra} note 8 at 180.

\textsuperscript{35} Michael Moore, “Causation and Responsibility” 16:2 (1999) at 11. For a related example see \textit{Frye v. City of Detroit}, 256 Mich. 466, 239 N.W. 886 (1932) where a pedestrian was hit by an automobile, thrown into the path of a street car, and then struck again. His widow was denied recovery as she could not establish which impact killed him.
death had it not been for the second attacker. Both factors were necessary to bring about the specific death suffered by the victim.

This is essentially the same qualification Wright adds to the NESS test when he requires that only sufficient sets actually operating at the time of an occurrence be considered. Both assailant's actions are both necessary and sufficient to produce the loss suffered in the circumstances. It is only where a very general description of the events is used to describe an occurrence that a sufficiency test is required. Mackie makes this observation about the example of two independently operating assassins, the first who poisons, and the second who empties their victim's canteen. He writes, respecting the statement “X caused Y” in the example:

If Y is the fact that the traveller died of thirst, then the puncturing of the can both caused and was necessary in the circumstances for Y. If Y is the fact that the traveller died on this journey, then the puncturing of the can neither caused nor was necessary in the circumstances for Y.

All “but for” tests necessarily and implicitly require, to some degree, limiting the generality of counterfactuals to the specific set of facts which actually occur. While a judge might parochially ask whether a death would have occurred, but for some factor, this question is clearly more limited than the language of the question suggests. All mortals die. If factor X does not kill person A, an infinite amount of other things might and at least one thing inevitably will.

36 A modified sine qua non test requires we consider an element's “but for” role only with respect to the specific occurrence—or, with respect to its role in the set of conditions actually operating at the time of the occurrence.

37 Mackie, supra note 3 at 46.
General and Specific Descriptions

There is, on the other hand, a great deal of flexibility with respect to the generality or specificity used when defining an occurrence. At its highest point of specificity, a sequence of events will have both been necessary to produce the exact set of circumstances which took place and, moreover, the only sufficient sequence which could have produced those circumstances. Nothing besides the life of Elvis Presley, for example, could have produced the exact circumstances of his death. As the description of the event becomes more generalized other factors appear sufficient to produce the generally-described type of loss.

Where actors operate concurrently to produce a generally-described loss the NESS test can be used to assign a causal status to sufficient, though unnecessary factors.

Suppose, for example, two children, without planning or coordination, simultaneously swipe at and strike a large Caesar salad, consequently knocking it off a kitchen table. If we describe the occurrence simply as a bowl of salad being knocked off a table, neither child's swipe is necessary for the occurrence. If one hadn't struck the salad bowl, the other would have. Both of the swipes, on the other hand, are independently sufficient to knock the bowl off the table. The swipes are, thus, NESS but not sine qua non causes. If, on the other hand, the description of the occurrence includes the exact pattern of the lettuce and croutons as they land on the kitchen floor, both swipes are sine qua non factors. Both are necessary for sending the bowl flying of the table with its precise velocity and trajectory and thus are necessary to the arrangement of salad on the floor.

38 “Strict sufficiency”, Wright, Pruning, supra note 8 at 1020.
39 Becht and Miller similarly write respecting the fires example: “in minute detail, it would probably
Wright offers two arguments against the idea that the generality employed in a description affects the necessity of contributing factors and, thus, the relevance of the NESS test. He argues that it is circular to, first, describe an event in detail (“as it came about”) and then to suggest that the details were necessary to the event, so described. Wright’s argument, however, demonstrates the same circularity. Wright suggests that irrelevant circumstances which happen to contextualize a cause-effect relationship can be discounted as “causes” if we limit ourselves to considering the sufficient sets actually operating in the circumstances. But how do we determine whether an factor is actually operating? Wright provides no guidance. He writes, referring to the “desert traveller” example:

How this result came about is the very issue to be resolved. To include how the result came about in the description of the result is to assume an answer to the causal question before it is posed. Thus, in the desert traveller example, the question of whether the traveller died by thirst, by poisoning, by both, or by neither, is precisely the issue that the causal inquirer is supposed to resolve. To describe the death as ‘death by thirst’ is to assume an answer to that question before it is posed...

He suggests that assuming, as Mackie does, that the victim “died by thirst” rather than simply “died” decides the answer to the causal question before it is asked. This is only partly true. We might infer that the victim “died by thirst” when we discover a dehydrated body. This link in the causal chain is assumed. However, that is not what is in issue. It is simply part of

appear that the defendant's fire was a course, for the positions of the smoke, ashes, and some parts of the ruins might well have been caused by the defendant's fire.” In A. Becht and F Miller, The Test of Factual Causation in Negligence and Strict Liability Cases, at 18, quoted by R. Wright, “Causation in Tort Law”, supra note 8 at 1779. See also, Cole, who argues that the degree of specificity in describing the “counterfactual” employed in a “but for” test affects the viability of the test. Where a highly specified counterfactual is employed, there is limited empirical basis to test the validity of the counterfactual assertion and the “but for” test becomes more speculative. “Windfall and Probability: A Study of ‘Cause’ in Negligence Law: Part I. Uses of Causal Language” California Law Review 52:3 (Aug 1964) 459.

Wright, Pruning, supra note 8 at 1025.

Ibid.

Ibid.
the narrative. The causal puzzle is whether the first or second assassin played a causal role in that particular death. The question is “Why was the victim's body dehydrated?” or “Was the conduct of one or both of the actors necessary in bringing this about?”. Where a house is destroyed by fire and we want to know which of two possible fires caused the destruction, we can assume that the house was *destroyed by fire* and not simply *destroyed*. Where a horse is spooked when motorcycles pass, we assume it was the noise which spooked the horse rather than some other factor and from there try to figure out which motorcycle was responsible. Mackie's point is simply that these assumptions can be made in varying degrees of generalization. It is only *broad descriptions* which yield the problem of *overdetermination*.

Second, Wright argues that using *specific descriptions* of an event makes every circumstance or factor situating an event a “cause” and does not allow us to distinguish between causes and correlations. Wright uses the example of two shooters whose bullets simultaneously strike a victim's heart. Wright asks us to imagine that a third party was climbing Mount Everest at the time. He argues that employing a specific description of the heart wound leaves it impossible to distinguish between causal and correlative factors situating the attack such as the third party's ascent of Everest:

> None of the three conditions by itself was a but-for-cause of the victim's death, but the cluster composed of all three of them was a but-for-cause. The aggregate but-for test does not distinguish the actual causes-- the two bullets-- from completely irrelevant conditions such as B's climbing Mt. Everest.

42 Mackie's argument was that if the injury is specifically described, the volley of both bullets together represents the “but 3 at 47; Wright, *Pruning, supra* note 8 at 1027. Wright makes the same argument in the article, “Causation in Tort Law” in response to a similar observation to Mackie's by Ronald Perkins, *supra*, note 8 at 1778.

43
It is true that employing a description which specifies every circumstance situating an event leaves it impossible to separate causal from merely correlative factors. As suggested above, every factor will be “but for” necessary to produce that precise description. On the other hand, Wright overstates Mackie's point. There is a clear middle ground between employing a perfect and universal description that includes every circumstance of an event and one that is hopelessly vague. There is clearly, for example, a middle ground between asking “Why did all the circumstances on the earth appear as they did at the moment of this victim's death, including B's climbing of Everest?” and asking “Why do people die?” We might ask why the victim died at that time. Posing the question in this manner leads to overdetermined causation, given that the two sufficient bullets pierced the victim's chest at the same time. On the other hand, we might ask why the victim died of two bullet wounds. Here, both bullets are sine qua non necessary to the latter question. Both questions are legitimate. The first is slightly more general than the second. However, both questions exist somewhere between perfect specificity and perfect generality. Every causal question put to the courts is always framed in this middle ground.

The NESS test, however, has a fatal problem with broadly-described consecutive causes. If an event's duration is extended far enough, we may be required to label things as “causes” that are intuitively non-causal. Imagine the victim dies immediately upon being shot. The second assailant however, does not know this and, like before, stabs the by-now-deceased victim. If our test is sufficiency and we ask what factors were sufficient to cause the loss that day, the stabbing, as a sufficient factor, is a cause of the death. Clearly this is wrong. Wright, accordingly requires the qualification that a NESS factor be operating at the time of a loss—
or, that the time frame under consideration be stated with specificity and precision. If we state the time of death with precision, however, the *sine qua non* test is perfectly acceptable and the NESS test adds nothing. *Sine qua non* like the NESS test, can explain that the stabbing party's act was *necessary* to precipitate and, thus, cause the death of the victim in the first scenario. It was *unnecessary* to the causal sequence in the second scenario where the stabbing party attacks an already dead victim. The NESS test is relevant to circumstances where an event is described so generally as to allow for multiple *concurrently-operating* factors *sufficient* to a produce a loss. It does not assist with “preemptive causation” questions where sufficient factors occur one after the other.

**Indivisible Entities and Uncertain Causal Processes**

As I have argued, the degree of generality or specificity given to the description of an occurrence affects whether the *sine qua non* test will support our “causal intuitions” or whether a NESS test is required. If a description is broad enough to include more than one sufficient causal factor or set of factors, certain factors will be redundant, unnecessary and thus, not *sine qua non* “causes”. In such circumstances, the NESS test can be used to assign a causal role to the redundant factors. This raises a further question. If an *sine qua non* test provides an adequate explanation of causation where descriptions of events are specified with precision, why would the law ever employ generalized descriptions? I suggest three reasons. First, the law must often assign responsibility for legally or putatively indivisible losses or injuries. Second, lawyers and judges often need to deal with minute or intricately complex causal processes unsuited for explication in the court room. Where it is intuitively clear that a factor contributed to a loss, but impossible, due to the complexity of a causal process, to
state the nature of that contribution, the law can employ a sufficiency test to ensure that a
causal contributor is held liable. The third reason borrows from the other two. Where a
reaction results after some threshold is breached, and more than one party is responsible for
the breach, a sufficiency test, again, can be used to hold parties accountable.

*Indivisible Consequences*

Let's slightly modify an example from above. Two independently-acting arsonists, both
unaware of the presence of the other party, start a house on fire. One arsonist starts a fire at
the north side of the house; the other, from the south side. Both fires are sufficiently
powerful to destroy the house in its entirety. Each fire, however, consumes one part of the
house and the two join in centre to destroy the remains. Both arsonists are held jointly and
severally liable for the entire damage.\(^4^4\) This result, in one sense, seems curious or fictive.
While both fires are responsible for the destruction of the entire house, the first fire destroys
one part of the house and the other fire, another. The fires only destroy part of the house
together. A more specific (and accurate) description of the damage would detail the portions
of the house destroyed by each.

Nevertheless, the house may be treated, for the purposes of law, as an indivisible entity and
not as the aggregate of its constituent parts. If the law were to simply hold the parties to
account for damage to the constituent parts of a thing, the “but for” test works fine. We ask,

\(^{4^4}\) Anderson v. Minneapolis, St. P. & S. Ste. M. Ry. 146 Minn. 430, 179 N.W. 45 (1920); Kingston v.
Chicago & NW Ry. (1927) 191 Wis. 610, 211 NW 913.
“but for this factor, would this part have been destroyed?” On the other hand, where the law holds parties to account for a thing as a whole the “but for” test leads to causal overdetermination. Both fires were sufficient to destroy the fire and, thus, neither was necessary. Here, the NESS test is useful. This accords with our general observation that broader descriptions of events can allow for multiple sufficient causes and, thus, create situations of causal overdetermination resolvable using the NESS test.

It is, thus, the conceptual or legal existence of a “house” as either a complete entity or a collection of its parts that dictates the procedure used to identify causes. Wright, although far from admitting that the NESS test is something to be used as a matter of policy or exigency, seems, on the other hand, to partly concede the point. He quotes Betch and Miller’s comments on the redundant fires example:

[I]n minute detail, it would probably appear that the defendant’s fire was a cause, for the positions of the smoke, ashes, and some parts of the ruins might well have been caused by the defendant’s fire.

Wright responds, as follows:

Again this assertion begs the question. Under this approach the defendant’s fire was a but-for cause only of the precise “positions of the smoke, ashes, and some parts of the ruins,” not of the destruction of the house, for which these precise details are irrelevant. The detailed description will be useful only if the plaintiff wishes to recover for damages that would not have occurred but for the precise position of the smoke and debris—a most unlikely lawsuit.

It is, thus, the pleadings, rather than the facts, themselves, that determine the applicability of

45 A. Becht and F Miller, The Test of Factual Causation in Negligence and Strict Liability Cases, at 18 supra, note39.
46 Wright, “Causation in Tort Law”, supra note 8 at 1779.
the NESS test.

*Indivisible Causes*

It would likely be very difficult, however, for a court to determine the precise measure of damage caused by each fire in the above example. It would certainly be so if the two fires joined prior to reaching the house and consumed the entirety of the house together. When the fires meet, the energy of one fire becomes an inseparable part of the energy of the other. Pollution is another important example. In situations where pollutants mix together, one source becomes impossibly mixed with another and determining the individual impact of either becomes impossible. In such circumstances, it is intuitively clear that both sources contribute to the injury. Each source forms an inseparable part of the same injuring force. Where the law is unable to distinguish between the contributions of one party to an injuring force and those of another, issues of *causal overdetermination* may arise. If two polluters

---


48 There is a need, however, to make an easily-glossed-over distinction between two types of uncertainty in cases which deal with multiple negligent parties. First, often there is a singular injury which clearly was the consequence of a single source, but the law is unable, due to the negligence of more than one party, to determine the responsible party. In such cases, there is an uncertainty as to the source of the injury. There are a number of famous examples: Two hunters go grouse hunting together knowing that there are other hunters in the area. As some grouse fly into the air, both hunters simultaneously take a shot. A nearby hunter is shot in the face. It is clear that the victimized hunter was shot by one of the two. It is, however, impossible to determine who was responsible. This example is similar to the grouse example noted above except that in this example it is unclear which bullet struck the victim. In the above example (of *causal overdetermination*) both bullets strike the victim at the same time. The present example reflects the facts of *Cook v. Lewis*, [1951] S.C.R. 830 (S.C.C.); *Summers v. Tice* (1948) 33 Cal. 2D 80. As another example, a woman is negligently hit by a bus, which then flees the scene of the accident. There are two bus companies in town and it is, thus, clear, that one of the bus companies was vicariously liable for her injuries. It is, on the other hand, impossible to say which of the two was responsible. *Smith v. Rapid Transit Inc.* 317 Mass. 469 (1945). A number of drug manufacturers sell the same drug. The drug is later found to cause medical complications. In the circumstances, there is no real way for any party injured by the drug to identify which manufacturer produced the dose that he or she was injured by. *Palsgraf v. Long Island Railroad Co.*, 162 N.E. 99 (N.Y. 1928). Notably, however, neither the “but for” nor the NESS test is equipped to deal with situations where an injury results from a single, albeit uncertain, source. If it is clear that one bullet struck the victim, but unclear which hunter was responsible, we cannot say with certainty that either bullet necessary or sufficient in the circumstances to
each dump a sufficient amount of pollutants into a body of water to cause a plaintiff injury, neither act is “but for” necessary for the injury. Yet both clearly contribute. NESS causation has relevance here.

Asymmetrical Overdetermination

A related problem of indivisibility arises in cases evidencing what Michael Moore has described as asymmetrical overdetermination. A causal process is asymmetrically overdetermined if one of the contributing factors plays a lesser causal role and is, in itself, insufficient to produce a loss, while another plays a larger, independently sufficient, role. Moore uses the fires example:

[S]uppose the fire set by the defendant is much smaller than the second fire; the two join as before and the resultant fire destroys the structure. The second fire would have been sufficient by itself to have destroyed the structure, but the defendant's small fire would not have been, since it would have been extinguished by the available equipment before it could have destroyed the structure.

The causal ramifications are, again, different depending on whether we employ a rich or a broad description of the loss. If we employ a rich description detailing the exact measure and magnitude of the injuring force and the exact nature and scale of the loss suffered, all of the constituent elements, both big and small, will be collectively necessary and sufficient. The amount of pollutants released by one of the lesser contributing parties forms a necessary part of the nature and scope of the injury and the level of toxicity created by the conduct of

50 Ibid.
all the parties. The “but for” test aptly handles this situation. If a \textit{broad description} of the loss is used, in contrast, the lesser factor will not be considered a cause on the application of \textit{either} the NESS or “but for” test. If the type of damage suffered from exposure to pollution can be described in a broad way, analogous to the destruction of a “house” as a singular totality, the release of a small, insufficient amount of pollutants at the same time as a large, sufficient amount will be neither necessary nor sufficient to produce the resulting loss.

Wright’s answer to the problem of \textit{asymmetrical overdetermination} is curious. On the one hand, Wright argues a \textit{rich description} yielding all (or more of) the causes attending an event \textit{necessary} is “useless tautology”.\footnote{Wright, \textit{Pruning}, supra note 8 at 1025.} He also, however, makes use of \textit{rich description} when stating that an independently insufficient element is an NESS cause. Referring to the concurrent fires example, he writes:\footnote{Wright, “Causation in Tort Law”, \textit{supra} note 8 at 1793.}

The same causal situation exists even if there were only two fires, one of which was independently sufficient and the other of which was not. The first fire was clearly a cause, since it was independently sufficient. But the second fire also was a cause. It was necessary for the sufficiency of a set of actual antecedent conditions which included another fire (the first) that was \textit{at least} large enough to be sufficient for the injury if it merged with a fire the size of the second fire.” The sufficiency of this set is not affected by the fact that the first fire was \textit{so} large that it would have been sufficient by itself... The word of the quoted condition, “\textit{at least} large enough,” is not a verbal gimmick. The condition is an actual one that existed on the particular occasion.

Wright’s answer is confusing. He argues that the condition “\textit{at least} large enough” qualifies the hypothetical sufficient set under consideration to those \textit{actually existing at the time}. However, if the larger fire was \textit{at least large enough} to destroy the house, the second fire is
unnecessary to do so. It is not “but for” necessary, nor is it a necessary element in a broadly
described sufficient set. It is true that the second fire is part of the conditions that actually
existed at the time of the occurrence. However, if the second fire was an unnecessary part of
those circumstances, it appears as a correlative rather than a causal factor.

Suppose one of the arsonists wears a red wristwatch or that the fire is lit under a full moon.
These also form part of the conditions which contextualize the destruction of the house. The
causal distinction between the smaller fire and the full moon or red watch is only explicable
by reference to the consequences wrought by the energy of that fire. A fire, no matter the
size, burns. Watches and moons do not. The smaller fire, moreover, played some role in the
burning of the house. It was not necessary to destroy the “house” as an indivisible unit. It
was, on the other hand, a necessary factor in effecting the nature, magnitude and shape of the
destruction. The moon and wrist watch had no impact on the nature or scale of the
destruction. It is only when we employ a rich description of the fire that the smaller fire
become a necessary element.

Perhaps this is what Wright is alluding to when he cryptically writes:

The word of the quoted condition, 'at least large enough,' is not a verbal gimmick. The condition is an
actual one that existed on the particular occasion.

The “actual one”, “at least large enough”, in Wright's comment, refers to an injuring force
effecting a very specific injury. It is the force which destroys an entity in a particular manner
and magnitude. The description of the “cause”, again, becomes inextricably linked to the
description of the “effect”. A smaller fire is unnecessary to effect the damage, broadly-
described, but was a necessary element in the richly-described destruction of the house.
Employing this *rich description*, however, leads us to Betch and Miller’s conclusion that:

in minute detail, it would probably appear that the defendant’s [smaller] fire was a cause, for the positions of the smoke, ashes, and some parts of the ruins might well have been caused by the [smaller] fire.

Again, Wright’s resistance to Mackie and Betch and Miller’s observation on this point seems misplaced given the necessity of using *rich descriptions* in order to make a smaller, independently-insufficient contributor a *necessary element* in a NESS (or equally, a “but for”) set.

Notably, in cases of *asymmetrical overdetermination*—or, where there is a contributing sufficient factor and a contributing insufficient factor—an NESS assessment never adds anything to *sine qua non* causation. If a *broad description* is used, the larger fire will be both necessary and sufficient for the loss and the smaller fire, unnecessary and insufficient. If a *rich description* is used, both fires are *sine qua non* necessary. The NESS test adds nothing. It only becomes useful in situations where a *broad description* is employed and where there is more than one sufficient factor (or set of factors). These cases are likely easier for our “causal intuitions”. If two fires large enough to destroy a house combine, it is easy to intuit that both fires “materially contributed” to the damage. The NESS test provides a sophisticated explanation of why this is so. The relevance of the test, on the other hand, is dependent upon the how the consequences are described. Many things can burn a house down. Only a particular set can burn a house down in a particular way. The relevance of the NESS approach is dependent upon whether a legal inquiry is looking to resolve a general or

---

53 *Supra* note 39.
particular causal question.

**Thresholds**

This conversation has an important application to causal questions that involve thresholds. Thresholds can have natural and physiological or social and legal bases. For example, after reaching a threshold of 80 milligrams of alcohol for 100 millilitres of blood, a person may no longer legally operate a motor vehicle.\(^5\) The rationale behind the threshold is that after reaching this limit an average person will no longer have the capacity to safely manage their driving. While the effects of alcohol on a human body is a physiological process, the threshold, itself, is social. It has been set in light of the danger that attends combing alcohol with a social activity. On the other hand, there are also strictly physiological thresholds. After a person reaches a blood alcohol level of 400 milligrams of alcohol for 100 millilitres of blood, he or she runs a serious risk of death. Death is a physiological threshold. Upon death, the heart will stop pumping blood and electrical activity in the brain will permanently cease. Thresholds can be breached by incremental contributions from a number of different sources. The (legal or physiological) injury incurred, on the other hand, is indivisible.

Where contributing factors are independently insufficient, but cumulatively sufficient to cause a loss, they are “but for” necessary for a threshold's breach. Suppose, for example, a patron at a bar is served alcohol by two different servers taking back-to-back shifts. The first server brings the patron a very large number of drinks sufficient to cause his death. He drinks these very fast and orders another round immediately from a second server, unaware

of the tab the patron ran up with the first. This round is sufficient, again, in itself, to put him past the point of death. If we think of the two rounds as having been consumed consecutively, the first is a necessary and sufficient cause of the death, and the second is not. A sine qua non test will suffice. As the first round was sufficient to cause his death, the second round was unnecessary. If we think of them as having been consumed concurrently, on the other hand, neither round is necessary. If he did not take the first round, the second would kill him and vice versa. An NESS analysis is needed.

In situations of asymmetrical overdetermination, again, sine qua non causation will suffice. Suppose the first round of drinks was more modest and insufficient to cause the patron much harm. Before these drinks are absorbed into the patron blood stream, the second server brings a round sufficient, in itself, to kill him. Was the first round a necessary and sufficient element in the death? If we simply consider whether the round contributed the patron's state of drunkenness or toxicity, the answer is clearly 'yes'. Intoxication is gradated. A person passes into a state of drunkenness in gradual or imperceptible degrees. The smaller, first round would serve as a necessary element in moving the patron towards that state. Death, on the other hand, is not a matter of degree. Its all-or-nothing nature means that a smaller, insufficient factor will be an unnecessary attending circumstance where death is succeeded by a larger, independently sufficient injuring force. The first round is, thus, an unnecessary part of the set of circumstances which caused the death and not an sine qua non or NESS cause.
Omissions raise a particular challenge for *sine qua non* causation. Even when the causal status of an omission seems intuitively clear, in any particular instance an innumerable number of other factors might have been capable of preventing the loss or injury suffered. Take the example referenced above. A garage does not repair the brakes on a car as required. The driver of the car approaches an intersection but omits to use the brake pedal. She strikes the car ahead of her. From a non-legal or mechanical point of view, describing either party's omission as a cause is problematic. The car was propelled to the scene of the accident by fuel and by its engine. It collided with the vehicle in front of it as a result of this force. It is certainty possible to state that the collision would not have occurred if the brakes had been operable and had been applied. However, the collision also would not have occurred if the driver diverted the vehicle up a runway truck ramp or if she simply didn't bother to drive that day, or if the vehicle was struck by lighting or was slowed to a halt by a pool of maple syrup. The number of things which didn't but might have occurred to stop the cars from colliding is limitless. They are, on the other hand, entirely speculative. It is, rather, the duties which bear on the occasion that make the omissions visible as causes. These duties cannot be garnered simply from the factual sequence of events. If the driver had no duty to stop, her failure not to do so would not be a “cause” of the accident—or would only be a “cause” in the most arbitrary and speculative sense. The question of duty must be answered prior to the question of cause and the two are inextricably linked.  

55 This conflation of a factual causation with policy has led some commentators to suggest that omissions are not properly referred to as “causes” at all. See, for example, Michael Moore, “Causation and Responsibility”, *supra* note 35 at 43.
On the other hand, once a legal duty is established, an omission can be analyzed much in the same way as factual causes. An omission tends to have a location in time, much like a factual cause. Its causal significance can be analyzed in relation to the rest of a factual sequence. If a parent omits to strap his child into a car seat and is subsequently involved in an accident and the child is injured, we can ask “but for the parent's omission would the child have been injured?”. Interestingly, however, unlike situations involving “factual causes”, where an incident occurs and there are multiple concurrently operating omissions, the NESS test is unnecessary. If any of the omitted acts had been performed in the circumstances, the incident would not have occurred. The omissions are sine qua non necessary to the occurrence. Suppose both of the child's parents had an independent legal duty to ensure the child was strapped into the car seat. Had either of the parents strapped the child in, the child would have been protected. Both (temporally concurrent and independently sufficient) omissions are thus sine qua non causes.

**Conclusion**

This chapter has argued that:

- The NESS test provides a useful tool for sorting out causation issues where:
  - an event is described broadly such that there are multiple sufficient causes;
  - these causes operate concurrently; and,
  - they form an indivisible injuring force or inflict an indivisible injury such as breaching a threshold.

- The NESS test does not add anything to a sine qua non analysis where:
• only one of the causes is sufficient (situations of *asymmetrical overdetermination*);
• the injuring forces take place consecutively (situations of *preemptive causation*);
• a *rich description* of the *injuring force* or the *injury* is employed such that all of the constituent factors become necessary to complete the description; or
• where there are omissions which are independently sufficient to cause the loss.

Both the “but for” and NESS test provide relatively clear and, in most circumstances, easily and predictably applied analytical tools for assessing whether a factor causally contributed to an outcome. On the other hand, these tools, themselves, far from determine the causation issue. As noted, the “but for” test can lead to the counterintuitive result that two factors operating concurrently and which are independently sufficient to produce a loss (and, thus, unnecessary) are not “causes” of the loss. In such circumstances, the NESS test is useful to support the intuition that both factors are, on the basis of their sufficiency, “causes”. The test, itself, is clear. An uncertainty exists, however, as to the manner in which the facts situating an event are constructed. The depth given to the description of an injury is something entirely independent of the heuristics used to assess causation. Our causal judgements are affected by these descriptions and this very human process of objectification and translation.56

Wright is correct to point out that plaintiffs generally do not maintain claims for precise details of a loss, but rather for broadly defined legal injuries. Minute details are generally absent in, or irrelevant to the facts and causes of action plead. The pleadings formulate a particular kind of causal question. As this chapter has argued, the formulation of the question, in part, determines its answer. This argument sits uneasily with the notion that we share universal “causal intuitions” capable of being objectively explicated in legal discourse. It, moreover, suggests that any particular causal analysis—including the NESS test—is going to be insufficient for every purpose. The relevance of NESS causation, rather, depends upon how we formulate our questions. While Wright's contribution to the literature is significant, his claim to have found the panacea to resolve “the causal issues in the problematic causation cases that have plagued tort scholars for centuries” is overstated.
Hart and Honoré's Normality Standard

A host of necessary factors attends every causal sequence. A comprehensive explanation of a causal process would need to explain the role of each of the operating factors. The law is not concerned, on the other hand, with providing comprehensive causal explanations. Rather, it is concerned, in the words of H&H, with attribute causation—"or with attributing responsibility for an occurrence to one of the causal factors or partitioning responsibility between a few. An attributive judgement, unlike a simply explanatory one, requires that these attending factors be narrowed to the relevant proximate or responsible causes of an event. The literature is largely divided between, in H&H's language, the "causal minimalists" who attach a minor role to the issue of causation and believe that factors are narrowed as matters of policy or exigency and the "causal maximalists" who suggest the concept of causation is broad enough to dispose of, if not the entirety of the liability question, a good deal of it.58

Take the following example. A car is travelling greatly in excess of the speed limit behind an ice cream truck. The driver of the truck spots some kids on the side of the road and slows down to make a sale. When the ice cream truck stops, the trailing driver is left with insufficient time and space to stop his vehicle and collides into the rear of the truck causing damage. There are potentially innumerable factors which are necessary to this occurrence. Some factors might include, for example:

- the presence of the ice cream truck;
- the presence of the automobile;

58 H&H coin the term in the preface to the second edition, ibid, at lxxiv.
• the presence of the children;
• the negligence of the defendant;
• the lawful stop of the ice cream truck;
• the existence of the defendant and the plaintiff;
• the existence of the parties' parents and grandparents;
• a gas attendant's act of filling the up the defendant's car with fuel;
• the fact that the defendant's car has an engine and wheels; or
• the invention of the wheel.

It is clear, however, that judges do not consider whether any particular automobile accident would have occurred but for the invention of the wheel or the conception of a defendant by his or her parents. The legal inquiry is much more focused. The legal inquiry is much more focused. There is no need for judges to provide a comprehensive picture of every antecedent which brings parties to a dispute into court. This would be impossible. Every antecedent is preceded by its own set of necessary causes and any given causal sequence undoubtedly stretches back in time ad infinitum. The

59 Ibid. There may be some relatively less contentious ways of limiting an incident's necessary or sufficient factors. First, for a thing to be considered a cause it must have an independent existence from the occurrence. In Mackie's words: "'X caused Y' presupposes that X and Y are distinct events". Mackie, Cement, supra note 3 at 32. It is superfluous to suggest that if a woman 'had never married she would not have been a widow'. H&H, "Causation in Law", supra note 4 at 114. This is by definition true. This argument might be uneasily extended to limit factors which are necessarily a part of every factor or occurrence of that type. Every human inherits their DNA from a pair of biological parents. We do not need to consider the fact that a defendant's parents existed and procreated at some point in history as one of the causes of an occurrence. We do not need to consider the presence of oxygen as a cause of a fire, since every fire necessarily demands oxygen. These are truisms that appear trivial or redundant to include on our list of causes. It is, however, not entirely certain that it is safe to exclude such general factors in all cases. As noted above an occurrence can be stated in varying degrees of generality. General causal laws are of more significance the more general the occurrence. If we are concerned with what caused fires, given its most general description, the necessity of oxygen is of obvious importance. If we are concerned with what caused a richly-described fire, on the other hand, the observation that oxygen was a necessary factor appears redundant or trivial. The explanation of a richly-described fire, rather, requires attention to the particular way the fire was started and developed. H&H, moreover, suggest that in instances where the presence of a factor generally productive of an occurrence is not expected, that factor appears as a cause.
factors a court takes into consideration are, at some point, “chopped off” from their history.\footnote{Bertrand Russell famously argued that this is marked by an analytically unacceptable arbitrariness. Bertrand Russell, \textit{Mysticism and Logic}, reprint, (Spokesman Books, 2007) at 175.}

Whether this “chopping off” represents a policy choice or, rather, something to do with causation itself, divides “minimalists” and “maximalists”.

H&H, as “maximalists”, maintain that the “chopping off” and “narrowing” processes have something to do with causation—or causation as is applied in our common sense judgements. They argue that causation is commonly understood by analogy to situations where an active element directly intervenes to bring about some change in the natural world that had been, up until that point, dormant and stable.\footnote{There are, however, entire classes of cases to which the “active force” metaphor has no bearing. H&H devote significant attention to causal relationships not captured by the simple metaphor of “active force” such as those pertaining to omissions, the creation of opportunity to cause loss, and acts of inducement. Such areas require distinct interpretive principles. H&H, “Causation in Law”, \textit{supra} note 4 at 51-59.} “Causes”, thus, become metaphorically invested with “forces” which “operate” and have “power”. The “static conditions” which situate an event, on the other hand, become the mere “backdrop” to the active “cause”.\footnote{\textit{Ibid} at 28-32.} Two general themes emerge from the metaphor. First, a “cause” is understood as that which disturbs the normal course of things. It represents an abnormality and deviation from the normal course. The metaphor can be extended from natural to social life. Where a course of conduct upsets the regular cadence of social life any attending losses are attributed to the responsible agent.\footnote{\textit{Ibid} at 33-41.} Second, “causes” become understood, by virtue of the metaphor, as the product of deliberate human agency. Once a voluntary human act is identified within a causal sequence we would
seldom trace preceding events back further in time.\(^{64}\)

This chapter argues that the use of abnormality and normality to distinguish “causes” from “conditions” takes on two shades of meaning. On the one hand, “normality” is understood as a strictly empirical measure. This is how H&H understand or relate the concept of “normality”. On the other, “normality” is something which “normalizes” and has “normative” import. While H&H attempt to distance their theory from a *normative normality*, this version clearly rests uneasily with its empirical counterpart in the same linguistic and conceptual space. The distinction between *empirical normality* and *normative normality* might be seen to very loosely reflect the distinction in the text between abnormality which is assessable against a static backdrop of custom or convention and abnormality assessable against the normal progress of “chains of causation”. I argue, however, that both of these standards have very dramatic effects on substantive legal outcomes and policy.

**Normality as Custom**

In the introductory chapters, H&H suggest that “causes” are distinguishable from “mere conditions” by their “abnormality” and conspicuous “operation” against the “background” or “medium” of the routines or affairs of ordinary life.\(^{65}\) It is, rather, the factor which “makes the difference” which is assigned the role of “cause”.\(^{66}\) Causes appear against the backdrop

\(^{64}\) *Ibid* at 68-81.

\(^{65}\) *Ibid* at 39.

\(^{66}\) *Ibid* at 35.
of “human habit, custom or convention”. A positive breach or omission to act in accordance with social mores or to fulfill a particular social duty, thus, will appear as the “cause” of an associated loss, where other “normal” factors hold. While this concept of abnormality is outlined in a few short pages, it informs the text throughout as a conceptual limit on the potentially innumerable sine qua non (or NESS) factors.

**Normality and Sine Qua Non**

The text is unclear, however, how abnormality and sine qua non interrelate in the day to day mechanics of judicial decision-making. Specifically, it is unclear whether the concept of abnormality informs the decision-making process as a limiter of preselected sine qua non factors or whether the concept informs the process of selecting the sine qua non factors themselves. The lack of clarity on this point reflects, more generally, some uncertainty as to what standard should serve as our sine qua non comparator when we consider what might have transpired had a test factor not occurred.

Richard Wright provides an exacting way to isolate the relevant causes-in-fact. He suggests the only factor a judge need consider when engaging in sine qua non (or NESS) reasoning is the tortious element of the defendant’s conduct. The focus is narrowed precisely to the

---

67 Ibid at 37.
68 Ibid at 38.
69 Robert Cole, for one, argues that it is not intuitively clear how a counterfactual world should be structured in order to separate the sine qua non (or NESS factors) from mere correlative factors, that there is no a priori standard that best serves this purpose and, thus, the process of sine qua non reasoning is indeterminate. Cole, “Part II”, supra note 9.
70 Wright, “Causation in Tort Law”, supra note 8 at 1806-7. See also, for example, S. Shavell, “An Analysis of Causation and the Scope of Liability in the Law of Torts” Journal of Legal Studies, 9:463, where Shavell suggests that causation-in-fact should be assessed by determining the consequences of a defendant not
causal implications of a defendant's failure to meet his or her standard of care. All the other factors which contributed to an occurrence are simply beyond the judicial pale. To do so, a judge compares the defendant's negligent conduct to a hypothetical (“counterfactual”) situation where the defendant met his or her minimum standard of care. This test provides a focused and predictable procedure for determining legal causation.

On the other hand, this process is limited to resolving issues of causation put into question by legal pleadings. His theory can go no further. Imagine, for example, that a defendant is involved in a collision travelling at a speed of 100km/h. Imagine that the speed limit and minimum standard of care in the circumstances is 80km/h. Under Wright's theory, we simply ask whether the accident would have occurred if the defendant was travelling at the speed limit. Now imagine the speed limit is, instead, 100km/h and that the defendant is travelling the speed limit at the time of the collision and has met her standard of care. In this situation we have no comparator to use in our counterfactual and, thus, are unable to consider the causal significance of driving at that speed. The consequences arising from non-negligent conduct are causally inexplicable. While this procedure is acceptable, it will be confined in relevance to the narrow issue of causation before the court. Wright is not able to provide an altering his or her level of care.

71 Wright develops this argument in the paper “Causation in Tort Law” in reply to Robert Cole's criticism of the use of counterfactual experiments as being necessarily vested with policy concerns. Cole discusses the example of the child who darts out between two parked cars into the road and is struck by a speeding driver. Cole acknowledges that the sine qua non test and the use of a counterfactual world where the driver's speeding did not occur is a relatively settled legal test for determining the connection between the defendant's speed and the accident. Cole, however, suggests that we are faced with an infinite number of choices in the counterfactual world we construct. We might compare what happened to a world where the driver drove the speed limit or, where the driver simply diverted his course earlier in her trip. We might construct a world where the driver drove at 1 km/h less than she did, in fact, or 2 km/h less, or 2 km/h more. The possibilities are endless. Cole suggests that the way that the construction of counterfactual worlds are manipulable enough to allow the causal question to either inculpate or exculpate the defendant. Robert Cole, “Windfall and Probability: A Study of 'Cause' in Negligence Law. Part II.”, supra note 9.
explanation of legal decision-making that is reconcilable with our “causal intuitions” or with causal judgements made outside of the law more broadly.

H&H's project is more ambitious. They attempt to explicate a theory which broadly reflects the way causation is understood in our common sense and “the ordinary man's stock of general notions”. The H&H text does not comment, on the other hand, on how counterfactual worlds are to be constructed or whether or not this involves the standards of normality and voluntariness. Three possibilities seem to emerge.

As a first possibility, *normalcy* serves a counterfactual comparator. A test factor—for example, the defendant's tortious conduct—is measured against conduct *normally* expected of the defendant in the circumstances. If, in a world where the defendant behaved *normally*, the plaintiff's injury would not have occurred, that factor is deemed *sine qua non* necessary. If the loss would have occurred regardless as to whether the defendant was acting normally or not, the factor is not a *sine qua non* cause. The same can be said of non-human factors. If, for example, a farmer's crop is destroyed during an unusually intense period of rainfall, we can speculate as to whether the crop would have been destroyed if rainfall had been *normal*. This procedure is sufficient to keep a ubiquitous, generally-prevalent “background” factor off of our list of causes. The conception of a defendant by his or her parents could not be

72 H&H, “Causation in Law”, supra note 4 at 91 [emphasis added]. H&H recognize that the concept of “common sense” is inherently less certain than the rigorous proofs demanded in the tradition of analytical philosophy. They suggest, however, that perhaps method ought to shift “to examine and chart in some detail the actual use made in given disciplines or in ordinary discourse of the key expressions like 'cause', 'consequence', and 'effects'”. The suggestion of a shift in the philosophic method to account for discursive uses of concepts shows the influence of Wittgensteinian and Austinian linguistic philosophy on H.L.A. Hart at the time he and Tony Honoré were writing Causation in Law.” See, P. Goodrich “The Role of Linguistics in Legal Analysis”, *Modern Law Review*, 47:523 (1984).
considered as a “cause” of a automobile accident, for example. This factor is normally (in fact, always) present and, as such, we would be left without a comparator for our counterfactual. Accordingly, it would not appear on our causal radar.

A second possibility is to employ the concept of abnormality only after the *sine qua non* factors have been selected. The determination of responsible or proximate causes occurs, in two (or, perhaps, three) steps. First, *sine qua non* variables are selected by a procedure independent of the concept of normalcy. Second, the *sine qua non* causes are narrowed. The factors which appear as exceptions to the ordinary course of things are identified as “causes” and the last such intervention (or set of sufficient interventions) as the proximate cause(s). A third step might include narrowing these factors further based upon any outstanding policy concerns. Both the ice cream truck driver's lawful stop and the trailing driver's speed, in the example mentioned above, are *sine qua non* necessary. The second step considers which of the two represents the abnormal element. The speed stands out as abnormal and, thus, in H&H's theory as the “proximate” or “responsible” cause.

While the authors do not explicitly adopt this approach, the text hints in this direction. The authors describe the modern bifurcation of the cause-in-fact question and the narrowing process with qualified approval. They suggest that the bifurcation has “real merits”, but that it is misleading to oppose the two as “factual” and “non-factual”. Rather, in their words,

73 See Leon Green, supra note 8.
74 H&H, “Causation in Law”, supra note 4 at 110
75 Ibid at 111
the second half of the bifurcated question are not *inventions* of the law but mirror principles characterizing ordinary non-legal thought.76

Their theory of causation involves something more than *sine qua non* causation rather than simply a revision of that concept:77

We must distinguish between the mistaken claim that all that is meant by “cause” (apart from the contribution of legal rules or policy) is condition *sine qua non*, and the general principle that no event happens without a cause.

*Abnormality* and *voluntariness* complement rather than restate the test for *sine qua non* causation.

The bifurcation of the causal analysis, however, leaves the question as to the choice of comparator unanswered. We still need to know the standard to be employed in the counterfactual used to determine the necessity of possible causes-in-fact. One possibility is simply to consider the non-presence or non-existence of a factor. We could determine, for example, the necessity of the factors situating a car crash by removing the presence of either driver. While the presence, for example, of an innocuously passing pedestrian is not likely necessary for the production of an accident, the presence of both drivers will be. We then narrow this list, further, using the concept of abnormality. If the defendant was negligent and the plaintiff guiltless, it is the defendant's conduct which appears as an exception to the ordinary course of things and, thus, as the responsible cause of the accident. While this method will yield results which seem intuitively correct in many cases, it will not in others. Take Cole's example. A defendant is speeding. A child darts out between parked cars and is

76 Ibid at 110.
77 Ibid at 129.
struck by the defendant's car.\textsuperscript{78} Suppose that, had the defendant been driving the speed limit, the accident still would have occurred. Here, use of non-presence in a counterfactual is clearly unacceptable. The conduct of the defendant and of the child both rank, where non-presence is our comparator, as \textit{sine qua non} necessary. From this list, the defendant's speed appears as abnormal. His conduct, thus, represents both a \textit{sine qua non} cause of the accident and an abnormality and, accordingly, a responsible cause of the accident. This cannot be correct.

In a third possibility, the first half of a bifurcated analysis follows Wright and the second half, H&H. To determine the \textit{sine qua non} relevance of the defendant's conduct we simply construct a counterfactual where we remove the \textit{tortious aspect of the defendant's conduct} and replace it with conduct which minimally meets the standard of care required in the situation. A speeding defendant who would have struck a darting child notwithstanding his speeding is, thus, not a \textit{sine qua non} contributor. If the defendant's tortious conduct had been a \textit{sine qua non} cause, we would then determine whether it fell outside of the ordinary course of things.

While this process is intuitively acceptable, it may be inconsistent with H&H's project for a couple of reasons. First, it is difficult to say what, if anything, the concept of abnormality adds. If a defendant breaches his standard of care, he will have presumably departed from the ordinary and normal course of things. Second, the use of the a minimum standard of care as a comparator in the counterfactual will only reveal the causal significance of things which

\textsuperscript{78} Cole, “Part II”, \textit{supra} note 9.
are in breach of such a standard. As noted above, this is an acceptable but narrow way to understand causation. It would be impossible, using this comparator to understand the causal significance of non-tortious human conduct or the causal significance of natural and non-human forces. In order to provide a broader view of causation reconcilable with non-legal thought another comparator is required. It would, however, be inconsistent with H&H's project to apply one set of causal tools to explain the significance of negligent conduct and another for all contributing factors.\(^{79}\)

Of these three possibilities, the first is likely the most reconcilable with H&H's project. The use of non-presence as a counterfactual comparator is clearly problematic. The use of a minimum standard of care leaves the theory unable to reconcile non-legal with legal causation. The first answer, on the other hand, allows for broad causal judgements that, arguably, accord with common sense, intuitive and non-legal causal understandings. Both negligent and non-negligent as well as human and non-human acts or elements can be considered against the common standard of normalcy.

**Normality and Normalization**

This usage, however, involves determining causation, as a putatively factual issue, on the basis of an actor's normalcy. The attribution of legal responsibility becomes undoubtedly linked to judgements about what is considered normal and what is considered abnormal. This raises a number of concerns. For one, the concept of normality is amorphous and, thus, subject to manipulation by lawyers and judges. Normality might simply operate as a guise

\(^{79}\) *Ibid* at 110.
for the interests of the litigants to any given dispute. The imprecision in the concept, moreover, raises concerns as to whether H&H's causation can serve as a predictable basis for effective dispute resolution. H&H are mindful of the theory's limits. They recognize that a core of common sense is bordered by a “penumbra” of uncertainty in hard or novel cases. They insist, however, that the core is clear enough to broadly inform a wide range of legal problems and that hard cases have:

never deterred English law from treating questions which involve them with a process for determining causation in law as questions of fact, and distinguishing them from general questions of expediency, justice, or social policy.

Even if we were satisfied that a normalcy standard was sufficiently predictable, the use of normality in attributive judgements raises serious concerns about the effect such a standard has on substantive legal outcomes. The use of normalcy as a measure of human conduct is a relatively recent social phenomenon. The concept of human normality in its modern manifestation is an 18th century innovation, most notably connected with the writings of Adolphe Quetelet and his concept of the “Average Man”. A large literature chronicles how statistical concepts of human normality have become increasingly prevalent over the last two centuries. H&H's invocation of normality is wedded to this particular social location. This is not overly problematic to their theory's integrity. H&H's concern is to give substance to

---

80 Ibid at 26, 92.  
81 Ibid at 92.  
common sense and intuitive principles which guide non-legal causal judgements. They do not claim an a priori knowledge of causation, but are rather concerned with the, albeit contextual and shifting, causal knowledge employed by the “ordinary man”, by “the historian, or by “the lawyer”, for example. On the other hand, this necessarily involves normality in governance. The objectification of norms involves fixing units of comparison and quantifying diversity along common metrics. It involves theorizing human objectives and rationalizing choice in light of those aims. It divides the abnormal from the normal. It requires, but also transmits, a tremendous ideational power and technical apparatus. An exploration into “norms as governance” literature is beyond the scope of this comment. However, it is fair to say that the invocation of normality exposes H&H to this sort of criticism. It accordingly requires putting their separation of factual from policy-based causal judgements under closer scrutiny.

H&H do not specifically address the normative implications of using a normality standard in attributive judgements. They do devote a fair amount of attention to distancing factual from instrumental causation. They insist that their use of abnormality to distinguish causes from conditions is not simply “dictated by our practical interest”. They recognize, on the other hand, that the element which appears, in common sense, as a “cause”, is often that which humans are able to manipulate or control and, thus, that the questions of policy and fact are closely linked. A doctor is said to rightly employ different causal understandings of a

86 See, for example, Michel Foucault, History of Sexuality: The Care of the Self, (Vintage Books, 1990).
87 They directly address this argument, H&H, supra note 4 at 36
88 Ibid.
patient’s medical reaction than the lay observer since her concern differs from that of the lay person. The medical profession's understanding of normality develops to address particular problems in medicine. H&H regard the connection between “controllability” or “policy” and causation, on the other hand, as indirect and unnecessary:

A gardener whose duty it is to water the flowers fails to do so and in consequence they die... It is not merely a breach of duty on his part, but also a deviation from a system or routine. It is, however, true that in such cases there is a coincidence of a deviation from a usual routine with a reprehensible dereliction from duty.

This answer is acceptable enough, in itself. However, the fact that people employ common sense judgements is not sufficient to elevate common sense judgements to status of fact. Common sense creates social facts, facts which are contextually dependent and facts which carry the world view and privilege of the party responsible for objectifying them. H&H recognize the difference between how causes are understood in science and by lay people and in other fields of knowledge. H&H make no attempt to ground the worlds in which we pass and exchange causal concepts to some objective reality. They are simply assessable in themselves. But these worlds are filled with politics. If we acknowledge that common sense and normality are political creatures, it makes it difficult to accept their division between factual and instrumental causation theories. It does not mean that the invocation of normality is necessarily the same thing as the invocation of policy objectives. Causes, perhaps, appear when a loss results from a breach of human “habit, custom or convention” whether or not we

89 Where the lay person, in H&H's example, suffering from indigestion may see his having eaten parsnips as the “cause” of his problem a doctor might point to the man's ulcerated condition. Ibid 36-37.
90 Ibid at 38 [emphasis added].
91 Ibid at 11-12.
have any interest in maintaining those habits or conventions. Nevertheless, attributive
judgements made against the standard of habit and custom clearly reinforce and perpetuate, if
not restate and define them.

**Normality as Ordinary Consequences**

*Original and Intervening Causes*

Another arguably distinct concept of “normality” is employed side by side in the text with
the notion of normality-as-custom.92 “Normality” is also used to refer to the regular and
expected way in which a causal sequence will run its course. A “cause” disturbs or interrupts
the normal dynamics of a sequential process or “chain of causation”. It represents an
abnormality because it changes the trajectory of this preexisting sequence. A normal
sequence might normally inflict a certain degree of damage to a party or the public or simply
exhaust itself, harmlessly. A “cause” interferes with this process. H&H write93

> If a contingency is, on account of its abnormality, to negative causal connection it must be an event and
> one later in time than, or possibly simultaneous with, the wrongful act. But a state of the person or
> thing affected existing at the time of the wrongful act (a 'circumstance'), however abnormal, does not
> negative causal connection.

This type of normality is secondary to or contingent upon passive or static normality. It only
arises when the peace of the static world is already breached by an original cause.

“Normality” encompasses the way in which people and things react to or interact with an

92 The dual usage of the term was noted by Wright in “The Nightmare and the Noble Dream: Causation
(Oxford University Press, 2008) at 172. See also, Patrick Kelley, “Causation and Justice: A Comment”
*Washington University Law Quarterly* 635 (1978) at 638.

93 H&H, supra note 4 at 172 [emphasis theirs]. H&H, rather, regard the relevance of the situating factors as
being limited to the assessment of damages rather than to the determination of the causation question itself
at 176.
original breach.

H&H suggest the normality of both original and intervening causes can be determined using a similar analysis:94

The defence of the principle is that it is rooted in common-sense ideas of human action and causation, which picture human beings as intervening on a stage already set. The action consists in the difference made to the scene, and ends when a new actor appears on an abnormal event disturbs the ordinary course of things.

They suggest a number of ways a causal chain can be severed by an intervening act or event. Most importantly, “the free, deliberate and informed act or omission of a human being, intended to exploit the situation created by defendant, negatives causal connection”.95 A human actor who interferes with the manner in which a causal sequence normally exhausts itself, effectively, takes ownership of the sequence and is held responsible for any resulting consequences.96

“Causes” might be said to have both natural as well as social consequences. The natural consequence of the rain, for example, in some circumstances, will be a rainbow. A “causal”

94 Ibid at 179 [emphasis added]. Patrick Kelley writes: “Hart and Honore's argument plays on the ambiguity in the phrase, “normal course of events”. In describing any purely physical system, the “normal” course of events is the expected course of events, at least theoretically predictable by reference to general rules of physical events... In describing human activity, the “normal” or expected course of events refers to traditional or habitual behaviour.” “Causation and Justice: A Comment” Washington University Law Quarterly, 635 (1978) at 638.

95 H&H, “Causation in Law”, supra note 4 at 136 [emphasis theirs].

96 Similarly, if an abnormal interjecting physical element (“acts of God”) or an abnormal conjunction of otherwise normal events (“coincidences”) displaces the ordinary course of a causal sequence, such events are deemed responsible causes. The text reads: “The basic principle is that normal physical events, even subsequent to the wrongful act, do not relieve a wrongdoer of responsibility but that an abnormal conjunction of events (in this case the wrongful act and the third factor) negatives causal connection, provided that the conjunction is not designed by human agency.” Ibid at 162-163.
event carries a number of natural consequences which flow from the event in the ordinary
course. An event will also, on the other hand, have social consequences. When faced with
an abnormal interference with the regular cadence of social life, people, as H&H argue,
respond in relatively patterned and predictable ways. There are likely normal ways people
manage or mitigate or, on the other hand, heighten or exacerbate the effects of an event. One
way of understanding intervening causes is to treat natural and social consequences alike.
The determination of cause can be understood as an empirical matter. The normal response
serves as a comparator. Those who divert from the normal response are liable.

Normality, Forseeability and Natural and Probable Results

The use of “normality” as an empirical standard, in some ways, reflects the “natural and
probable results” or “foreseeability” tests employed in law. H&H, however, distinguish
their normality standard from both. H&H deal with the “natural and probable results”
standard summarily:

A reaction which is both natural and probable may negative causal connection. For a man who has
suffered a serious injury owing to a defendant's negligence it may be natural to decide to take a holiday,
and it may be probable that he will do so, but it does not follow that he can charge the tortfeasor with the
cost of it.

H&H devote more attention to distinguishing their approach from foreseeability, though they
do so in a similar manner. While foreseeability is regarded as a “shortcut” to the results

97 See, for example, Milwaukee and St. Paul Railway Company v. Kellogg (1877) 94 US 469 and
Palsgraf v. Long Island Railroad Co., 162 N.E. 99 (N.Y. 1928), respectively.
98 H&H, supra note 4 at 158 [emphasis added].
99 In earlier parts of the text, H&H use the “normality” and “foreseeability” standards interchangeably. A
rendered through a “normality” test, they view the two as distinct. Many things, they argue, are unforeseeable from the perspective of the defendant and, yet, proceed as normal consequences of an occurrence:  

...what is unforeseeable to the defendant before he has done the negligent act may well become foreseeable at a later stage, after the initial harm has occurred, on the basis of the fuller information then available.

Some unforeseeable losses occur quite in line with the normal order of things. Suppose an unwitting tourist arrives in a particularly windy locale, unaware of the normal weather for the season. She lights a fire during a period of calm. When the winds pick up the fire spreads and causes loss. The tourist did not foresee and, perhaps, could not have reasonably foreseen that a torrent of wind would have stirred and spread the fire, despite the fact that such winds occur in the normal course. Conversely, there may be losses which are foreseeable in advance, if as a slight probability, but which represent an abnormality to the ordinary course.  

The generalizations employed in a normality standard are broader than those used in a foreseeability test. They are based upon a social understanding of the ordinary course of things rather than an individually situated one. Moreover, foreseeability is assessed prior to the loss from the particular point in time that the defendant engaged in a wrongful act. A normality assessment determines the status of each sine qua non factor as they arise

footnote on page 163 of their text states this expressly: “For present purposes it does not matter if 'unforeseeable' is substituted for 'abnormal'.” Ibid.  

100 Ibid at 265.  

101 The Law and Economics literature, thus, substitutes the concept of an all-or-nothing foreseeability standard with that probability or likelihood of loss. See, for example, Steven Shavell, Foundations of Economic Analysis of Law, (London: Harvard University Press, 2004) at 5-6.
sequentially and not simply from the *ex ante* perspective of the actor.\(^\text{102}\) This is a considerable distinction and leads H&H to draw the conclusion that a foreseeability test conflates the question of “culpability and compensation in a single formula”\(^\text{103}\). The foreseeability test is, in other words, a policy limit masquerading as a causation heuristic.

It is not clear, however, why foreseeability judgements are policy-based where normality judgements are not or why foreseeability accords less with common sense. H&H argue that common sense assigns causation by reference to the metaphor of an active intervention in a static world. The concepts of normality and abnormality loosely reflect the dichotomy between active and passive. This dichotomy, however, is impossible to maintain when considering the status of intervening causes.\(^\text{104}\) In these instances, the static world is already breached by an original intervention. Our concern is whether the dynamics of the chain of events set in motion by this original active force become interrupted by another active force. The abnormality of an *intervening event* might be contrasted against a normal way of responding to a situation. It cannot, on the other hand, be contrasted against a passive “background” of “mere conditions”.

*Normalizing Contingencies*

If there is little suggesting why normality standards more aptly reflect common sense

\(^{102}\) H&H at 262.

\(^{103}\) Ibid.

\(^{104}\) The metaphor collapses, moreover, if we consider that a failure or omission to respond to a sequence of events is often considered as being causally connected with an abnormal or unforeseeable loss. If a plaintiff was hospitalized and subsequently died because the operating surgeon was reading magazines instead of attending to the surgery, we would assign mark this omission as a cause. H&H note that omissions generally operate as an exception to the active/passive metaphor, *Ibid* at 51-59.
than foreseeable standards and *foreseeability* is said to conflate policy issues with factual issues, we might question whether *normality* can be similarly criticized. Consider the “normalizing” effects of using normality standards to assess the progress of causal sequences. There may be an ordinary or regular way in which societies handle risks, mitigate injuries, or deal with emergencies or contingencies. The assessment of intervening events against a normality standard, thus, involves a “normalization” of social responses to dangerous or injurious forces. An abnormal intervention which exacerbates an injuring force, would be deemed the responsible cause of injuries in excess of “normal” injuries. Perhaps in many circumstances, the public or certain classes of the public are normally expected and/or have a duty to neutralize an injuring force or minimize the scope of its damage. Rescuers are held accountable for gross negligence in their rescue efforts, doctors for malpractice, fire fighters or police officers for the derogation of their duties, etc. In normal circumstances, these actors will attempt to interrupt an injurious chain of events. A deviation from the competent performance of these roles represents an abnormality and attracts liability. The application of a “normality” standard operates as a deterrent against the deviation from conventional responses to social contingencies. It documents and clarifies, as well as incentivizes and regularizes social responses.

This is the version of “normality” which H&H seek to avoid as it would seem to “bring culpability and compensation within a single formula”. H&H's use of “normality” as a heuristic of common sense, on the other hand, is difficult to distinguish from “normality” as a tool to normalize. The agent who fails to respond to a contingency according to
normal protocol stands out from the other *sine qua non* causes as the responsible cause, perhaps, as a simple matter of common sense. But also as a matter of social policy. Both versions of “normality” uneasily and inextricably share the same language.

**Conclusion**

H&H write: 105

...men have discovered that nature is not only harmful if we intervene, but is also sometimes harmful unless we intervene, and have developed customary techniques, procedures, and routines to counteract such harm. These have become a second 'nature' and so a 'second' norm.

This passage evidences both ways in which a “normality” standard might be employed or understood in law, either as an *empirical* measure or a *normative* standard. Normal procedures, policies and routines are developed in response to particular social problems. They represent “techniques” to control hazards and minimize harms or promote stability and the normal order of things. On the other hand, norms, in H&H's writing, become a “second 'nature'”. They crystallize into customs and take on a life of their own. And they are capable of being measured as such. In many circumstances, it may well be possible to base judgements upon, or act in accordance with norms without giving any consideration to their policy rationale, purpose, or to the role of the norm in community or social life.

105 H&H, *supra* note 4 at 37.
The invocation of empirical norms or norms-as-second-nature, without considering the underlying rationale for the norms or the effects of normalization, carries with it, however, some danger. Perhaps customs and conventions arise within traditions which privilege the interests of some and prejudice the interests of others. Without closer scrutiny, the use of these norms by lawyers and judges might perpetuate, concretize, reify or mystify these prejudices and give them the force of law. It is likely true that our conceptions of “common sense” are intensively political. The right to speak on behalf of “common sense” is likely subject to hard-fought ideological contests. Judges and lawyers themselves, moreover, also develop, formalize and perpetuate linguistic and structural habits and conventions. Perhaps the problems of litigants undergo a process of translation into a legal language that takes on a life of its own. It may be important to ensure, or at least aspire, to anchor the norms employed in attributive judgements to the values and concerns of parties to a dispute and the public at large. H&H anchor their theory of causation to norms operating in the community more broadly. We might, further, insist that the norms, themselves, be anchored and justified by their service to the communities in which they develop.

Robert Cole, for example, notes that H&H rely strictly on litigated scenarios and the usage of causal concepts by the legal profession as the source of their examples of common sense “normality”, “Part I” supra note 9 at 468.
Finding the “Normative” in the “Empirical”: Examples from Insurance Law

I have argued that the use of normality as an empirical benchmark in attributive judgements affects substantive outcomes. It “normalizes” and has “normative” implications. Both the empirical and normative versions of normality necessarily and uneasily share the same linguistic and conceptual space. The tension between the two is evident in the case law. This chapter explores the construction of the term “accident” in insurance coverage cases. In these cases courts must assess whether a loss was caused “accidentally” or not. If a loss was the product of an “accident”, it falls within the scope of insurance coverage. If not, the insured bears his or her own loss.

H&H's causation theory is hard to apply directly to the insurance cases. On the other hand, Canadian coverage decisions strongly evidence the tension in H&H's theory between employing a strictly empirical limiting device to reduce causes-in-fact to the proximate or responsible cause of an occurrence and the development of substantive policy. The current test for “accident” approximates a foreseeability test. As will be discussed below, the current approach is to assess the subjective expectations of an insured. If an insured does not expect a loss, the occurrence is deemed an accident. Where an insured, on the other hand, “courts

107 The cases do not borrow language directly imported from H&H. Insurance would seem to represent a hard set of cases for their theory. Take, for example, a comprehensive third party liability policy. Coverage applies to acts or omissions which are both “tortious” and “accidental”. A tortfeasor will have acted abnormally to attract tort liability, in H&H's theory. Their behaviour stands apart from the “background” of “mere conditions” due to its abnormality. They are, accordingly a responsible cause of the loss. On the other hand, in order to uphold coverage, the judge must find that the event was “accidental”—or that an abnormal coincidence of events “caused” the loss. These are opposite findings. The same abnormality standard cannot apply to both the negligence and coverage decisions.
the risk” of loss, it is attributed to the insured. This type of causal analysis, like H&H's analysis of intervening abnormalities, concentrates on the chain of causation. While H&H study the causal significance of each event as they arise and an expectations assessment considers the sequence from the ex ante perspective of an insured risk-taker, both measure the progression of events against what is (individually or socially) expected to occur. On the other hand, courts are loathe to uphold coverage where an insured takes a reckless risk with little redeemable purpose. “Courting the risk” takes on another meaning. Where a party takes wanton risks with no redeemable purpose, his or her conduct is generally not covered. This has clear policy implications (not unlike those which attend the attribution of responsibility to parties who act abnormally). Risk-takers must justify their risk-taking. Risk-taking must have some redeeming merit or useful purpose. Superfluous or inexplicable risks are disincentivized. As this next chapter argues, these instrumental aims heavily influence the disposition of the cases. But not expressly.

The Definition of “Accident” in Canadian Coverage Cases and the Unspoken “Useful Purpose” Test

The coverage allowed by the courts under comprehensive “accidental loss” policies is considerably more liberal than it was 50 years ago. It affords policyholders a broad right to indemnity for almost everything but subjectively expected losses. This chapter argues that this liberality is constrained by a concern in the case law for the nature and character of

108 B.Billingsley, General Principles of Canadian Insurance Law, First Ed. (Markham, ON: LexisNexis, 2008).
an insured’s risk-taking. Where risks are taken superfluously or with no apparent redeeming value, courts are more likely to deny coverage. Where an insured can point to some redeeming value in the risk, courts are less likely to do so. I argue that the phrase “courting the risk”, 110 takes on two distinct meanings. First, at some point, the risks associated with an activity are said to be so substantial as to suggest that an insured expected the resulting losses. Such losses are deemed to have been “courted”. The second meaning, on the other hand, suggests that a party “courts the risk of loss” where risks are taken solely for the experience of risk-taking, in and of itself, and not for any other redeeming benefit. Policyholders who act for the mere “psychological gratification in living on the edge... or in order to impress others with their bravado”111 are said to have “courted” their losses.

The chapter, first, outlines the expectations test and its evolution in the case law. It considers the interpretation of the term “accident” in liability, property and accidental death cases as they tend to borrow definitions from and inform each other.112 Second, it contrasts the expectations approach to instances in the case law where courts show a concern for the utility of an insured’s risk-taking and argues that judges tacitly employ an unarticulated “useful purpose” test.

110 The decision Candler v. London & Lancashire Guarantee & Accident Co. of Canada et al (1963) 40 D.L.R. (2d) 408 at 421 is often quoted as authority for the proposition that where an insured “courts the risk” of loss, the policy will not pay.

111 Martin, supra note 109 at para 25.

112 See, for example, B.C. Master Blasters Inc. v. Aviva Insurance Co. of Canada 2006 BCSC 1488, 42 C.C.L.I. (4th) 190 at para 23.
The Expectations Approach

The Subjective Standard

In the 2003 decision, *Martin v. American International Assurance Life Co.*, the Supreme Court of Canada provided the present legal test to distinguish accidental losses from losses attributable to the conduct of an insured.\(^{113}\) The case involves a claim under an “accidental death” provision in a life insurance policy providing benefits in case of death by “accidental means”.\(^{114}\) The claimant was a family physician who became addicted to morphine and Demerol after an orthopaedic injury. The insured was on a physician-monitored gradual withdrawal program at the time of his death. He died of a self-administered intravenous overdose of Demerol.

The Court held that the question as to whether the death was the result of “accidental means” is to be ascertained by assessing the subjective expectations of the insured.\(^{115}\) If death was an unexpected consequence, coverage is upheld. If death was expected, coverage is denied. In many situations courts will not have direct evidence as to an insured’s expectations,

\(^{113}\) *Supra* note 109.

\(^{114}\) The insurer resisted the claim on the basis that an “accidental means” policy excludes losses which result from the consequence of an insured’s voluntary conduct, even if the results of that conduct were unanticipated. The Court rejected this argument and held, in the absence of strict exclusionary language expressly limiting coverage in situations where an insured voluntarily participates in the chain of events culminating in a loss, an “accidental means” policy covers the unintentional and unanticipated results of intentional acts. *Ibid* at paras 9 – 18. This decision succeeds a vigorous debate over the legal distinction between “accidental means” and “accidental results” policies. The distinction, rejected in *Martin*, had been previously upheld by the Supreme Court of Canada in *Smith v. British Pacific Life Insurance Co.* [1965] S.C.R. 434 and was upheld by the Alberta and British Columbia Courts of Appeal in the decisions *Leontowicz v. Seaboard Life Insurance Co.* (Alta. C.A.) and *Columbia Cellulose Co. v. Continental Casualty Co.* (B.C. C.A.), respectively, and, famously, by the United States Supreme Court in *Landress v. Phoenix Mutual Life Insurance Co.* (U.S. S.C., 1933) (Justice Cardozo issued a much-cited dissent). The distinction was later rejected by the British Columbia Court of Appeal in *Ostrinski Estate v. UNUM Life Insurance Co. of America*, 2 B.C.L.R. (3d) 76, [1995] 4 W.W.R. 144, for example. The viability of the distinction between means and results policies is not directly addressed here.

\(^{115}\) *Supra* note 109 at para 21.
particularly where there is little evidence surrounding an insured’s death. In such instances, a subjective-objective assessment is required.\textsuperscript{116} Courts must ask whether a reasonable person in the position of the insured and with the insured’s particular attributes and understandings would have expected the loss. On the facts of the case, the Supreme Court held, given the insured’s professional background and experience with the drug, his optimism about his future and the compromising state in which he was found, that the insured had not expected to die. Accordingly, coverage was upheld.

Notably, at trial, the judge found that the insured had deliberately engaged in an activity that posed a \textit{high risk of death}.\textsuperscript{117} This did not determine the issue. The Supreme Court confirmed that an occurrence can be deemed “accidental” even if the insured acted with foresight of the possibility of loss.\textsuperscript{118}

The trial judge found that Dr. Easingwood was engaging in a “particularly hazardous” activity. Given this high risk, death “would not reasonably be viewed as an unlikely result”. This set the legal threshold too low. As this Court affirmed in \textit{Stats, supra}, death as a result of even highly dangerous activities may be accidental. The issue is not whether the activity was dangerous, or even whether death was likely, but whether the insured expected or intended to die. While the trial judge said the

\textsuperscript{116} \textit{Ibid.} The two-step approach reflects that set out in the seminal United States Court of Appeals for the First Circuit decision in \textit{Wickman v. Northwestern National Insurance Co.} 908 F.2d 1077 (1\textsuperscript{st} Cir. 1990).

\textsuperscript{117} The Supreme Court of Canada does not expressly reject this factual finding. On the other hand, in discussing the trial judge’s failure to consider evidence suggesting that the insured had not expected to die, the Court suggests that some of the evidence supports the conclusion that the insured had not even turned his mind to the possibility of death, \textit{supra} note 109 at para 37: “The first set of facts concerns the circumstances in which Dr. Easingwood’s body was found. The body was found in a disheveled state inappropriate for someone who anticipates death as a potential result of his actions. He was lying prone in his office with his glasses broken on the floor beside him, with his jeans partially pulled down, revealing the site where he had injected the Demerol. These facts point strongly to the conclusion that Dr. Easingwood did not expect to die; indeed they \textit{suggest that he did not so much as turn his mind to the possibility that death would result from his actions [emphasis added]}”. At any rate, the Court held, as discussed above, that the having foresight of the possibility of death is too low a test for “accident” and does not determine the issue.

\textsuperscript{118} \textit{Ibid} at para 34.
death could not be viewed as an unexpected event, he equated this with objective likelihood, not with whether the insured expected to die.

An insured’s deliberate participation in a “particularly hazardous” activity or foresight of the possibility of loss is not enough to show that he or she expected the loss. On the other hand, it is possible to expect something without intending it. In Martin, the insurer conceded in an agreed statement of facts that the insured had not intended to die. The Court held that an insured must more than foresee, but need not actually intend, a loss to be denied recovery:

The authorities clearly stipulate that the mere fact that someone has engaged in a dangerous or risky activity does not rule out the possibility that death was accidental, absenting special exclusion clauses in the insurance policy. However, the decision to “court the risk” of death, as Spence J. phrased it in Stats, becomes at some point equivalent to an intention to die (citations omitted).

The Court does not explain what, in addition to foresight, is required to establish an insured’s expectations. The decision simply suggests that in “this small but difficult class of cases, trial courts must work out the results as best they can”. It is clear that, at some point, the risks of foreseeable losses become so considerable that they allow the inference of an expectation to die. Where this occurs is unclear.

Objective Standard Cases

The expectations test has come to widely define the term “accident” in insurance policies

119 Ibid at para 33.
120 Ibid at para 23.
121 Ibid at para 24.
122 The same observation has been made of American case law on the definition of the term “accident”. See, for example, P.J. Daue “The Foolish Insured and Double Indemnity” Washington and Lee Law Review 20:143 (1963) at 147.
generally and has an applicability to both first-party and third-party policies.\textsuperscript{123} The decision in \textit{Martin}, however, represents an extension or evolution of preceding doctrine, rather than a marked departure. While the legal rule has evolved from a more restrictive objective standard to the more liberal subjective standard employed in \textit{Martin}, the jurisprudence has generally drawn its definitions of “accident” from an assessment of the expectations and intentions of the insured or from those of a reasonable person in the place of an insured. In the 1903 House of Lords decision, \textit{Fenton v. Thorley and Co. Ltd.}, Lord Lindely stated:\textsuperscript{124}

The word “accident” is not a technical legal term with a clearly defined meaning. Speaking generally, but with reference to legal liabilities, an accident means any unintended occurrence which produces hurt or loss.

The objective statement of this rule is offered, for example, in \textit{Halsbury}:\textsuperscript{125}

The idea of something haphazard is not, however, necessarily inherent in the word: it covers any unlooked for mishap or an untoward event which is not expected or designed, or any unexpected personal injury resulting from any unlooked for mishap or occurrence. The test of what is unexpected is whether the ordinary reasonable man would not have expected the occurrence, it being irrelevant that a person with expert knowledge, for example of medicine, would have regarded it as inevitable.

The objective test outlined in \textit{Halsbury} was adopted by a number of courts prior to the decision in \textit{Martin}.\textsuperscript{126} This test is obviously more restrictive. An insured holding an honest, though unreasonable, belief that his or her conduct would not cause a loss is denied cover.

\textsuperscript{123} See, for example, \textit{B.C. Master Blasters Inc. v. Aviva Insurance Co. of Canada} 2006 BCSC 1488, 42 C.C.L.I. (4th) 190 at para 23. James Rendall argues that the equation of the definitions of “accident” in third-party cases with those in “accidental death” claims occurred as early as \textit{Stats v. Mutual of Omaha Insurance Co} (S.C.C.) in “Drink, Drive and Die! Then Ask us to Define Accident” \textit{Manitoba Law Journal} 9:101 (1978-1979) at 110-111.

\textsuperscript{124} 1903 A.C. 443 at 453.


under the objective approach in circumstances where coverage would be upheld today.

On the other hand, the distance between the objective and subjective tests is not as marked as it might appear at first blush. For one, the distinction is of no importance where a court has little or no evidence of an insured’s state of mind. In such instances, as noted from *Martin*, a subjective-objective approach—or one considering the expectations of a reasonable person in the place of the insured—applies.

Second, courts in the past seldom went so far as to suggest that objective *foresight* or negligence was sufficient to negate coverage. In the decision *Marshall Wells of Canada v. Winnipeg Supply and Fuel Co.* an insured corporation was denied coverage under a third party liability policy after losses resulted from the collapse of a water tank. Agents of the corporation knew prior to the loss that the tank was inadequately supported. The majority of the Manitoba Court of Appeal found that the resulting losses were foreseeable, that the insured had voluntarily assumed a calculated risk and that this risk-taking took the occurrence outside of the definition of “accident”. In a dissenting judgement, Justice Freedman held that negligent conduct can be considered “accidental”:

> That a mishap might have been avoided by the exercise of greater care and diligence does not automatically take it out of the range of accident. Expressed another way, "negligence" and "accident" as here used are not mutually exclusive terms. They may co-exist.

Justice Freedman’s dissenting approach was adopted by the Supreme Court of Canada in

---

127 49 W.W.R. 664, 1964 CarswellMan 44.
Canadian Indemnity Co. v. Walkem Machinery & Equipment Ltd.\textsuperscript{129} While the Court maintained an objective standard for assessing expectations,\textsuperscript{130} the Court rejected the notion that evidence of calculated or negligent risk-taking was sufficient to disprove an event as “accidental”. The decision reads:\textsuperscript{131}

[On the insurer’s argument], the insured would be denied recovery if the occurrence is the result of a calculated risk or of a dangerous operation. Such a construction of the word “accident” is contrary to the very principle of insurance which is to protect against mishaps, risks and dangers.

In the decision Stats v. Mutual of Omaha Insurance Co., the Supreme Court of Canada extended the scope of “accident” coverage to gross negligence:\textsuperscript{132}

Therefore, I am in agreement with Blair J.A. when, in giving reasons, he said that there was every justification for the learned trial judge's description of the deceased woman's conduct as dangerous and grossly negligent but that that was far different from finding that the insured actually and voluntarily "looked for" or "courted" the risk of the collision that killed her.

The inclusion of negligence, calculated risk-taking and gross negligence within the sphere of activity that can be considered “accidental” allows a distinction between reasonable foresight of loss and the reasonable expectations of loss—a reasonable person can be held to have foreseen a loss without having expected it. This distinction reduces the seeming strictness of the reasonableness standard.

\textsuperscript{130} Ibid at para 19.
\textsuperscript{131} Ibid at para 15.
\textsuperscript{132} [1978] 2 S.C.R. 1153 at 1162.
Third, the Court in *Martin* held that an insured can be *deemed* to have subjectively expected his or her losses in situations where an insured’s risk-taking is particularly excessive—or where the insured “courts the risk of loss”.\textsuperscript{133} The Court, for example, cites with approval *Candler v London & Lancashire Guarantee & Accident Co.*\textsuperscript{134} where coverage was denied to the beneficiaries of an insured who fell to his death during a stunt performed for his friend. *Martin* reads:\textsuperscript{135}

One might speculate that the trial judge [in *Candler*] concluded that, despite a hope and belief that he would survive, the insured had knowingly adverted to the risk and must have, on some level, expected death.

This, again, closes the gap between the objective and subjective expectations test. The finding that a party acting without the intention to cause a loss actually “adverted to” or “courted the risk” of loss (the subjective standard) is not far removed from the finding that a party’s expectation to escape loss was unreasonable (the objective standard).

*Chain of Causation Cases*

A related body of older cases focused on whether or not a chain of events set in motion by an

\textsuperscript{133} Prior to *Martin*, Professor James Rendall argued that the decisions of the Supreme Court of Canada in *Stats v. Mutual of Omaha Insurance Co.* (Ont. C.A.), affirmed (S.C.C.) and, more particularly, *Straits Towing Ltd. v. Washington Iron Works* (S.C.C.) seemed to have eliminated the “courting the risk” category of cases. Rendall, James A. Annotation 2003 CarswellSask 191, 2003 SKQB 116. With respect, this conclusion seems overstated. The cases do adopt a more liberal standard. On the other hand, the Court in *Stats* expressly states that the expectations approach is consistent with the earlier decisions which applied the “courting the risk” standard at paras 20-21. See E.M. McVitty, “Foreseeability—The Scene of the Accident Revisited”, *Manitoba Law Journal*, 10: 443 (1979-1980) at 448-449. Moreover, the Court’s statement in *Martin* at para 26 that an insured who does not desire an outcome could, nevertheless, knowingly avert to the risk and, “on some level”, expect loss suggests the old category is alive and well.

\textsuperscript{134} *Supra* note 6.

\textsuperscript{135} *Supra* note 5 at para 26.
insured was interceded by an event falling outside of the normal course of things.\textsuperscript{136} The House of Lords decision in \textit{Claxton v. Travellers Ins. Co. of Hartford} offers a frequently cited definition:\textsuperscript{137}

'Accident' has been defined to be an unusual and unexpected result attending the performance of a usual and necessary act. It is an unexpected event which happens as by chance, or which does not take place according to the usual course of things. Any event which takes place without the foresight or expectation of the person affected by the event; or is an unusual effect of unknown cause and therefore not expected.

\textit{Couch's Cyclopedia of Insurance Law} is similar:\textsuperscript{138}

In other words, an accident is an undesigned contingency, a casualty, a happening by chance, something out of the usual course of things, unusual, fortuitous, not anticipated, and not naturally to be expected.

Some Canadian courts have used the standard of “natural or probable results” or the “usual course of things” to determine whether or not a voluntary human act was considered the proximate cause of an insured’s loss and, thus, outside of an “accidental loss” policy. In this analysis, where voluntary conduct is succeeded “in the ordinary course” and a loss results, the loss is deemed not to have been “accidental”. Where voluntary conduct is succeeded by “something out of the ordinary” and a loss results, the interceding element is considered the

\textsuperscript{136} James Rendall, for example, draws a similar distinction in older Canadian cases between those which equated “recklessness” with an intention to cause loss and those which analyzed the “proximate cause” of the loss. James Rendall, “Drink, Drive and Die! Then Ask us to Define Accident” \textit{Manitoba Law Journal} 9:101 (1978-1979) at 104-108.

\textsuperscript{137} at 485. The definition of an accident as an “unusual and unexpected result attending the performance of a usual and necessary act” traditionally defined, what has been called, an “accidental results” policy. In older cases, courts suggested that in an “accidental means” policy covers the results following an unusual and unexpected source. As noted above, this distinction was rejected in \textit{Martin v. American International Life Insurance Co}. Notably, \textit{Couch's Cyclopedia of Insurance Law}, vol. 5 at 3976 criticizes the application of the definition in \textit{Claxton} to “accidental means” policies.

\textsuperscript{138} \textit{Ibid}, Vol. 10, sec. 41:6
proximate cause and the loss is “accidental”. For example, in *Greenway v. Saskatchewan Government Insurance Office*, where an insured was involved in a chase with the police and rolled his car while attempting a high speed turn at an intersection, his injuries were deemed the “natural consequences” of his course of conduct and nothing “fortuitous” was said to have interceded his conduct and his losses.139

There are important distinctions between a “natural or probable results” or “usual course” standards and an “expectations” analysis.140 “Natural and probable results” employs an *ex post* and conceptually objective understanding of causal processes. “Expectations” assess the individually-situated understanding of causation from the *ex ante* perspective of the defendant. On the other hand, in most cases a “natural and probable results” analysis will approximate an “expectations” analysis. Where an outcome follows as the “natural or probable consequences” of an action, the losses are deemed to have been expected.141 Where a voluntary act is interceded by events outside the “usual course of things”, the consequences


140 This distinction has been the site of a great deal of debate in legal theory. The move from the use of “natural and probable results” standards to “foreseeability” standards in tort law, more generally, bears the influence of the American Realist movement and represents one of their most important and lasting achievements. Morton Horowitz, “The Doctrine of Objective Causation” in David Kairys ed., *The Politics of Law: A Progressive Critique* (New York: Pantheon Books, 1982).

141 *Crisp v. Delta Tile &Terrazzo Co.*. [1961] O.J. No. 170. [1961] O.W.N. 278, [1961] I.L.R. 1-029 (C.A.) at para 6. On the other hand, the cases considering whether losses which result from medical causes are said to have been produced by “external” means within the scope of coverage or “disease” or “internal processes” tacitly or explicitly excluded from coverage impose an additional stipulation: the event must be “external” as well as “unexpected”, as *per Martin*. See *Gibbens v. Co-operators Life Insurance Co.* 2008 BCCA 153 at para 21-22. When a voluntary act is attended by unexpected consequences, the proximate cause analysis can also be used to distinguish whether the resulting chain of events falls within coverage or not. If a consequent event that is deemed the “proximate” or “efficient” cause of the loss falls outside of coverage, for example, if it is the product of a “disease”, the losses will be excluded. See, for example, *Wang v. Metropolitan Life Insurance Co.* [2004] 72 O.R. (3d) 161 (Ont. C.A.). If, on the other hand, the “proximate” cause is deemed to have been “external” to the body of the insured and the consequent physical reactions occurring within the insured deemed “mere conditions” or “passive allies” in bringing about the loss, the loss is covered. See, for example, *Voisin v. Royal Insurance Co. of Canada*, 1988 CarswellOnt 716, 33 C.C.L.I. 1 at 24.
come “unexpectedly”. The two approaches are often used interchangeably. Consider, for example, the language of the decision in *MacIsaac v. CNA Assurance Co.*:142

In my view it is the unexpectedness of the result that is the essence of what is meant by the term "accident or accidental" in policies of this sort. If a result of the type or kind that actually happens could be foreseen as a natural and probable result of the act engaged in, then the actor can be said to be courting the risk. What follows then ceases to be accidental even though it was hoped that a particular result would not follow [emphasis added].

This line of cases imposes an objective standard. The “usual course” or “natural and probable results” standard references understandings of how events generally proceed, rather than how they would be expected to proceed in an insured’s understanding. In this sense, the older cases are inconsistent with *Martin*. On the other hand, a “usual course” standard might be modified to correspond to a subjective approach. The test in *Couch* might be adapted to read:

An accident is an undesigned contingency, a casualty, a happening by chance, something which is perceived by an insured as being out of the usual course of things, unusual, fortuitous, not anticipated, and not naturally to be expected.

If this qualification were added, the standard would more or less, if imperfectly, reflect *Martin*. An insured would likely have a hard time convincing a court that, while he or she appreciated that a loss would normally follow a certain act or invention, the insured did not expect the result in the particular instance at issue.

---

At any rate, under both approaches, at some point, a level or risk will be so high that a court will deem that the insured reasonably or subjectively expected the loss as a probable result. These tests, on the other hand, are not concerned with the social context of the risk-taking, the returns or redeeming benefits gained from the risk, or with the moral or economic implications of risk-taking. If the courts do, as this chapter argues, determine the scope of coverage by considering the reasonableness of risk-taking (to be distinguished from the reasonableness of an insured’s expectations of loss), the efficiency of risk-taking, or its propriety, for example, it does not occur doctrinally. The expectations test, whether subjective or objective, focuses merely on one dimension of the risk—the perceived likelihood of loss. It does not consider the scale of the risk or its benefit. These things, on the other hand, can dramatically influence the disposition of cases.

**The “Useful Purpose” Test**

Take the following two examples. First, an individual covered under an “accidental death” policy participates in a game of Russian roulette. The revolver has six chambers and only one bullet is chambered. The insured faces a 16.7% chance of death. A referee spins the cylinder and then, without ascertaining the position of the chambered round in relation to the firing pin, locks the barrel in place and hands the revolver to the insured. The insured pulls the trigger and is killed instantly. His beneficiaries claim against his insurer.

Second, an individual with the same coverage makes occasional trips to the Canadian north and is required to traverse the Mackenzie River at Fort Providence, NWT. In the summer, she can cross by ferry. In the winter, the river freezes over and automobiles are permitted to
travel over the river by way of an ice road. On April 13th the insured departs, expecting the road to be open. When she reaches the river, she finds it closed. Ferry service does not begin until around mid-May. She knows the road generally closes around mid-April and that the ice has yet to break up. She can be fined for crossing the closed road, but knows a number of people who have used it as late as the start of May, in some cases. It is very late into the night and the closest motel is 2 hours south and charges a mint. She estimates her odds of making it across at 83.3% and decides to chance the road. Halfway across the bridge, however, she hits a peculiarly weak patch of ice. The ice cracks, the insured’s truck falls under the ice shelf and the insured dies moments later. Her beneficiaries, likewise, claim against her insurer.

In both examples, the insureds face the same likelihood of death. Neither one desires it. They both intend to survive the risk. On a strict application of an expectations approach, there may be little to distinguish the two cases. The treatment of these examples by the courts, on the other hand, might well differ. The first is clearly excluded under an “accidental death” policy. The claim of the insured who dies playing Russian roulette was considered by the Court of Appeals of Georgia in the decision Thompson v. Prudential Insurance Co. of America143 and has been cited as a paradigmatic example of “courting the risk” of loss in many Canadian decisions.144

144 See, for example, Candler, supra note 110 at para 20; Martin, supra note 109 at para 25; MacIsaac, supra note 142 at para 23; Booth v. British Columbia Life & Casualty Co. 2003 BCSC 668 at para 27 (The court rejected the claim due to an exclusion for injuries incurred while voluntarily intoxicated. This ruling was upheld on appeal); Johnson v. Mutual of Omaha Insurance Co. [1982] O.J. No. 3543, 39 (2d) O.R. 559 at 49.
The outcome of the second example, on the other hand, is less certain. There are many instances where courts have upheld “accident” coverage for policyholders who participate in dangerous activities. It was considered an “accident”, for example, where a dentist died as the result of sniffing ether from a bathing cap to relieve stress and depression;\textsuperscript{145} where an insured died during autoerotic asphyxia;\textsuperscript{146} where a stuntman died while attempting to dive into a small pool from the ceiling of the Houston Astrodome;\textsuperscript{147} and where a hunter shot a member of his party after negligently firing into the woods believing that he had spotted a deer.\textsuperscript{148} The risk of loss was present in the mind of the insured in our example. The Supreme Court of Canada is clear, however, that “calculated risks” can still produce “accidents”. Risk-taking, even negligent or grossly negligent and calculated risk-taking, is not a bar in and of itself to coverage.

Perhaps the second example is most analogous to the facts of \textit{Trynor Construction Co. v. The Canadian Surety Co.}\textsuperscript{149} That decision involved the collapse of a bridge after an employee of the plaintiff misjudged the amount of weight that the bridge would be able to sustain. A senior driver with the plaintiff was tasked to deliver a bulldozer resting on a flat bed and attached to the back of a Mack tractor to the company’s gravel pit. On route, he came to a bridge. He stated to a co-employee: “I wonder how strong that is?” After considering the difficulty of lowering the tractor to the bank and driving across the stream bed, the decision

\textsuperscript{147} Soucek Estate 66 O.R. (2d) 140
\textsuperscript{149} (1970) 1 N.S.R. (2d) 599 (Nfld. C.A.).
was made to attempt the bridge. His testimony at trial was, as follows:\textsuperscript{150}

“So I asked him, I said, would you suppose that the bridge is strong enough to carry us? So he went down, he was gone quite some time, he came back up, and he said, 'It looks all right.' So we stood there, like he stood by the cab talking for a couple of minutes, we decided that there was no sign on the bridge or anything, so that we would try it. So I proceeded to cross the bridge. That's what happened.”

The insured’s losses were held to have been “accidental” and within the policy, notwithstanding that the driver took a “calculated risk” by trying to cross the bridge.

It is hard to say how a court would come down in our ice bridge example. The weight of authority, however, suggests strong arguments can be made in support of coverage. If these arguments prevailed, the distinction between the first and second examples is curious. If courts simply assess an insured’s perceptions of the \textit{ex ante} probability of loss, why would the court uphold coverage where an insured incurs a risk to avoid expenses when it denies coverage to the insured taking a comparable risk in Russian roulette? If there is a distinction to be made between the two instances, it does not have anything to do with the perceived likelihood of death. Rather, a distinction is found in the \textit{character} of the risk-taking and the redeeming upside \textit{benefit} of the risk. The \textit{expectations test} does not accommodate these considerations. These considerations do, on the other hand, colour the meaning of the phrase “courting the risk”. The phrase is used to refer to conduct that involves a substantial likelihood of loss.\textsuperscript{151} On the other hand, the phrase is also used to refer to those who, like

\textsuperscript{150} \textit{Ibid} at 604.

\textsuperscript{151} Professor Rendall divides this type of “courting the risk” cases into those in which an insured voluntarily exposes him or herself to danger and those in which an insured is reckless or wanton with his or her risk-taking. James Rendall, “Drink, Drive and Die! Then Ask us to Define Accident”, \textit{supra} note 123 at 110. With respect, this seems to be a fine distinction. As this chapter argues, the more salient distinction is between instances where an insured “courts the risk” by taking a risk with an extremely high probability of loss and
the participant in Russian roulette, engage in risks superfluously, gratuitously or for the purpose of the experience of risk, in and of itself.\textsuperscript{152} While the latter meaning is not reflected in the expectations test, it exerts a tacit influence on the case law.

\textit{Purpose and Credibility}

There is an obvious and uncontroversial way in which a concern for the character or purpose of an insured’s risk-taking influences the cases. There are many situations where the degree of foresight that an insured had into the possibility of loss is not clear. For example, where the circumstances surrounding the death of an insured under a life or accidental death policy give rise to suspicions of suicide it may be unclear whether, on the facts, the insured knew that his or her conduct carried the risk of death. In these situations, evidence of the purpose or motive helps a court make factual findings.\textsuperscript{153} The trier-of-fact might infer: “If the insured had no reason for taking the risk, he or she must have actually desired death and intended suicide.” This is a factual inference.

On the other hand, in many cases it is clear, either from direct evidence or because of the inherently risky nature of an activity, that an insured, while not desiring loss, had foresight into the possibility of loss.\textsuperscript{154} For example, in \textit{Candler} it was clear that the insured intended

\begin{flushright}
\textsuperscript{152} Sam Erman, for example, suggests American courts have not recognized acts of “high-risk bravura” as “accidental”. “Word Games: Raising and Resolving the Shortcomings in Accident-Insurance Doctrine That Autoerotic-Asphyxiation Cases Reveal”, \textit{Michigan Law Review} 103:2172 (2004-2005) at 2184.
\end{flushright}

\begin{flushright}
\end{flushright}

\begin{flushright}
\textsuperscript{154} The Court in \textit{Martin} describes this as “a small but difficult class of cases” in the context of “accidental death” claims, \textit{supra}, note 5 at para 24. On the other hand, this class makes up the lion’s share of third party
\end{flushright}
to take the risk of falling when prostrating himself on the coping of his balcony.\textsuperscript{155} In \textit{Walkem Machinery}, it was clear that the insured intended to run the risk of returning a self-unloading barge to the plaintiffs in a state of inadequate repair.\textsuperscript{156} In \textit{Trynor Construction}, it was clear that the insured’s employee knew that by attempting to drive the equipment across the bridge he ran the risk of it collapsing.\textsuperscript{157} In these cases, the question remaining is whether the insured’s perceived probability of loss is deemed “equivalent to the intention”\textsuperscript{158} to cause the loss. This is also considered a factual inference to be drawn from the circumstances.\textsuperscript{159} However, this is a different kind of inference than the one mentioned above. In the former case, the circumstances bear on whether the insured had \textit{foresight} and \textit{intended} the loss. In the latter, the circumstances bear on whether the insured’s \textit{foresight} equates with an \textit{expectation} of loss. The \textit{expectations} doctrine treats the two inferences that can be drawn from an insured’s purpose alike. An inference arises that the insured who acts with no seeming purpose probably desires and intends the loss (and, thus, “expected” it). An inference also arises that the insured who, while not intending loss, takes a calculated but superfluous risk, “courts” his or her loss (and thus “expects” it).
“Expectations”

The term “expected” is probably not the most fitting to deal with the second kind of inference. Is a loss any more expected if risks are taken for some purpose, than if they are taken purposelessly? The term “expectations” refers to the perceptions of probability or likelihood, rather than defensibility or rationality. Taking risks gratuitously or without reason does not make them any less probable or likely. A different test is required. This might be formalized as a “useful purpose” test.

American tests for recklessness in tort law might be candidates. *Prosser and Keaton on the Law of Torts* suggest an actor is reckless where:

the actor has intentionally done an act of an unreasonable character in disregard of a known or obvious risk that was so great as to make it highly probable that harm would follow, and which thus is usually accompanied by a conscious indifference to the consequences [emphasis added].

The *Restatement (Second) of Torts* defines recklessness as:

intentionally fail[ing] to do an act... knowing or having reason to know of facts which would lead a reasonable man to realize, not only that his conduct creates an unreasonable risk of physical harm to another, but also that such risk is substantially greater than that which is necessary to make his conduct negligent [emphasis added].

Both tests include an assessment of the reasonability of the actor’s risk-taking. At any rate, some test might be formulated such that, if an insured is unable to demonstrate some purpose

---


162 *Restatement (Second) of Torts* (American Law Institute,1965) § 500.
to his or her risk-taking, the claim fails. The distinction between the sort of utility derived from simple risk-seeking behaviour and utility carrying a “redeeming purpose” is likely a fine one and, in some circumstances, probably illusive. If such a test were formulated doctrinally it ought not, in my opinion, mask a moral disapprobation about the nature of an insured’s conduct.\textsuperscript{163} On the other hand, there may be relatively non-contentious categories of risks that people would commonly recognize as superfluous or as having no other purpose than the experience of the risk itself—the case of Russian roulette, for example. Courts might viably deny recovery to those whose conduct falls within such a category.

\textit{Examples from the Case Law}

A concern for superfluous risk-taking seldom creeps into legal standards. As an exception, take for example, the decision in \textit{Oakes v. Sun Life Insurance Co. of Canada}.\textsuperscript{164} In that decision, an insured was killed in a high speed chase with the police. The trial judge, in reasons delivered orally, found that the insured “courted the risk” of his loss and, as such, coverage was denied. The decision reads:\textsuperscript{165}

\begin{quote}
In this case I am convinced that Oakes was "\textit{actually and voluntarily looking for and courting the risk of the collision}" that killed him. It cannot matter, in my opinion, that the risk of being struck from behind was not as great as the risk of running into a southbound vehicle or of colliding with a police vehicle upon being forced over to the side of the highway. Oakes may and probably did expect to
\end{quote}

\begin{flushright}


\textsuperscript{165} \textit{Ibid} at para 6.
\end{flushright}
outrun his pursuers and to escape injury, but as in Candler, the collision which caused Oakes’ death was not “an unusual or unexpected incident associated with the deceased's actions [citations omitted; emphasis added].”

Notably, the trial judge distinguishes between the insured who “expects” to survive and the insured “voluntarily looking for” and “courting the risk” of death. Oakes “probably expected” to survive and yet was still “courting the risk of collision”. On the other hand, the decision curiously suggests at the end of the passage that Oakes’ death was not an “unexpected incident”. The reasons are scant, were delivered summarily and are too confused on the issue to form grounds for a legal distinction between “courted” and “unexpected” losses.166

Nevertheless, the distinction between expectations and courting the risk hinted at in Oakes is more supportable if we consider the disposition of the cases on their facts, rather than the legal standards invoked. In the frequently cited decision of Candler v. London & Lancashire Guarantee & Accident Co., the Court was faced with the claim of a man who placed himself on the balustrade of his 14th floor apartment in order to demonstrate his nerve to a friend.167 The insured lost his balance and fell to his death and his beneficiaries claimed on an “accidental death” policy. The trial judge found that the insured had courted the risk of falling and that, notwithstanding that the insured had “hoped and probably believed that he could accomplish the attempted feat without injury”, he still had “present in his mind the risk

166 James Rendall suggests that this case represents a “sighting” of the “undead” concept of “courting the risk”: supra note 135. As suggested in note 135, the weight of authority suggests that the expectations test and the courting the risk tests are compatible and complimentary. The decision in Oakes, on the other hand, demonstrates a conflict between the two. It may be, however, that the decision is simply not good law.

involved.” The lengthy reasons and detailed review of the law do not mention whether the character or benefit of risk-taking ought to inform the decision. On the other hand, the judge’s summary of the facts demonstrates a conspicuous concern with Mr. Candler’s character and the nature of his risk. Candler is described as a “very determined, egotistical and reckless individual”.\(^{169}\)

The judgement goes into considerable detail about the excessively useless nature of the risk. It compares this risk to other risky ventures from his life:\(^{170}\)

However, in most of the incidents referred to in the evidence, there was something to be accomplished by his unusual acts. In permitting an adversary to exert a dangerous hold on him in judo exercises he gained knowledge of the effect of such holds. His feat of climbing the mast of the boat in stormy weather was to make adjustments to the rigging. Riding a strange horse bareback over the hurdle was a means of testing out the horse in the contemplated purchase. Walking out on the catwalk surrounding the flagpole on the top of the hotel in Cleveland may have afforded him some thrill from the view of the city that was afforded thereby. His acts on the night in question in assuming the dangerous position he did on the top of the coping could have no useful purpose whatever except the obvious opportunity to convince Simmonds that he possessed sufficient nerve to accept the challenge that was associated therewith. His conduct was foolhardy and attended with the most obvious danger.

Unlike Candler’s other risks, the risk that killed him had no purpose. There was not “something to be accomplished” in the insured’s risk-taking and the claim failed.

\(^{168}\) Ibid at para 34.

\(^{169}\) Ibid at para 7.

\(^{170}\) Ibid at para 35.
The case contrasts with the decision of the Ontario District Court in *Soucek Estate v. Atlantic Mutual Life Assurance Co.*\(^{171}\) The deceased was insured under a standard “accidental death” policy. He was an experienced stuntman with a penchant for dramatic and death-defying acts. In 1984, for example, he locked himself in a specially-fitted barrel and was dropped over the Niagara Falls. The next year, he was, again, placed in a specially-fitted barrel, this time in front of thousands of onlookers at the Houston Astrodome, lifted 118 feet to the building’s ceiling and positioned over a small tank of water. When the rope was released something went wrong. Instead of falling into the tank, the barrel struck its side and Soucek suffered fatal injuries.\(^{172}\) His estate claimed on the “accidental death” policy. The Court upheld coverage. It found that the onus was on the insurer to exclude “stunts”.\(^ {173}\) It, moreover, held that, as Soucek had intended to survive the act, his death was accidental. The decision reads:\(^ {174}\)

Dealing with the facts of the case before me, it is clear that Soucek was a stuntman and one must assume that his intention was that the descending barrel would drop into the 12-foot wide tank containing 9 feet of water. In fact the foot of the barrel struck the side of the tank with great force.

This, to my mind, was the “accident” and as a result of that Soucek died.

The decision cites but does not expressly distinguish *Cander.* If there is a distinction to be drawn between the two it is likely not in the degree of danger facing the men. Both men were defying death with their stunts. The distinction, rather, is more readily explicable by reference to a “useful purpose” standard. As the court stresses, “Soucek was a stuntman”. He was chiefly employed to flaunt risk in front of thousands of onlookers. His stunt, like

\(^{171}\) *supra* note 147.

\(^{172}\) *Ibid* at para 1.


\(^{174}\) *Ibid* at para 14.
Candler’s, was a “test of nerve”. On the other hand, these tests were his business and his risk had a redeeming commercial value.

The British Columbia Court of Appeal considered the meaning of “accident” in relation to high-risk conduct in the decision, *C.J.A. v. American Home Assurance Co.* In *C.J.A.*, the insured died during auto-erotic asphyxiation. The insured placed a mask and bag over his face in order to reduce the flow of blood to the brain and heighten the sexual pleasure of masturbating. The insured was found to have miscalculated the amount of oxygen he required and as a result he died in the act. While the Court acknowledged that “the deliberate reduction of oxygen to one's brain is an inherently dangerous activity” it found that “the reasonable ordinary person, not affected by a moral judgement about the activity, would not see death as its objectively foreseeable result.” It seems curious and confusing that an insured participating in an inherently dangerous activity would be unable to reasonably foresee the possibility of death. The Court characterizes the event as a “misjudgement”. The characterization implies that the insured was, in fact, cognizant of

---


177 *Supra*, note 175 at para 10.

178 *Ibid* at para 6 (emphasis added).

the risks of death and calculated, if erroneously, the measures necessary to avoid death. The “objective foreseeability” standard is, moreover, inconsistent with the Supreme Court of Canada’s holding in Walkem Machinery and Stats that “calculated risk-taking” and “gross negligence” can fall within the definition of the term “accident”. The “objective foreseeability” standard is unhelpful.

The decision is less confusing if we consider the court’s assessment of the insured’s intentions. It reads: “the intended result of autoerotic asphyxiation is sexual pleasure”. Notably, the Court does not invoke the insured’s intention to survive (though this might be implied). Rather, it invokes the insured’s purpose of taking the risk—namely, sexual pleasure. It is the purpose that explains the event and distinguishes it from the useless, foolhardy class of dangers in Candler. The Court’s example is telling:

We are reminded of this regularly by those concerned with the safety of children who may find old refrigerators or bags made of plastic film attractive as toys...

…But if plastic film a child chose to put to his or her face were not removed or the door of the refrigerator in which a child chose to hide was not opened, I have no doubt all would agree his or her death was accidental.

The intention of the child, here, is clear—the refrigerator forms part of a child’s game. This purpose explains the risk and why death comes “unexpectedly”. Imagine, on the other hand, an who adult climbs into a refrigerator, not in a game of hide-and-seek, but rather as a

----------------------

180 Ibid.

181 The judgement in Bertalan Estate is similar. The Court concludes: “Common experience would suggest a suicidal intention. But that inference is negatived by the experience, knowledge, and motive of the insured.” [emphasis added], supra note 175 at para 6. The insured’s voluntary inhalation of nitrous oxide was explained by the insured’s pleasure-seeking and pattern of abuse.

182 A.C.J., supra note 72 at paras 10-11.
demonstration to her friends of how much oxygen she could deprive herself without suffering death. That degree of risk might be comparable to the insured’s in this case. Their expectations of survival would be the same. It is the purposelessness of the latter, however, which seems to render it outside the scope of “accidental death” coverage.

_C.J.A._ stands in contrast to _Davis v. Clarica Life Insurance Co._\textsuperscript{183} The insured was a young man known to the people that he roomed with to be a drug addict. The toxicologist’s report indicated a methadone overdose. His family claimed the “accidental death” benefits. The insurer did not argue suicide or self-inflicted injury.\textsuperscript{184} At trial, however, the claimants were unable to provide evidence of the insured’s intentions in the moments surrounding his death. The decision reads:

> The problem we have here is that as Ms. Kraft has pointed out, nobody was called who had been with Jamie before his death or at his death. All we know is that he was found dead and there is some indication that the people he was living with saw him lying on a mattress at two o'clock in the morning and phoned the police at eight o'clock in the morning. Particularly, we do not know anything about what was on his mind in the day or two before. Whether there was anything that would point to a wish to die or an intent to die or knowledge that if he took methadone he might die. Nothing about what he knew of the risks of the drugs. Some indication from unnamed person spoken to at the scene that said he was a drug user. Obviously unreliable evidence.

The trial judge found that the claimants had not discharged their onus to show that the death was “accidental” and, thus, within the scope of the policy and, accordingly, dismissed the claim.

\textsuperscript{183} 2004 CarswellOnt 9743.

\textsuperscript{184} _Ibid_ at para 15.
It is interesting, however, to consider the onus that seems to have faced the claimants.\textsuperscript{185} As the judgement indicates, there was no evidence tendered as to whether the insured wished to die or not. Where an insured takes a large and indiscriminate amount of drugs, an intention to cause self-harm or suicide might be inferred.\textsuperscript{186} Here, the insured evidently took enough methadone to kill himself. On the other hand, the insurer did not argue self-harm or suicide. Where there is no evidence of an insured’s subjective expectations, the court is entitled to consider what the expectations of a reasonable person in the place of an insured would have been. One might reasonably assume that if a person does not intend to die from some risk, he or she intends to survive it. This logic was applied, for example, by the British Columbia Court of Appeal in the decision \textit{Bertalan Estate v. American Home Assurance Co.},\textsuperscript{187} where the insured died from asphyxi-ation in the course of abusing nitrous oxide. It reads:\textsuperscript{188}

\begin{quote}
In the absence of an intention to commit suicide, the insured must have either miscalculated the amount by which he could reduce the oxygen flow to his brain or inadvertently did not pay sufficient attention to it.
\end{quote}

What is missing in the evidence in \textit{Davis}, however, is what purpose the insured had in taking his risk. The judgement seems to suggest that it is not sufficient for a claimant to simply rely on the fact that there was no record of self-harm or evidence of suicide. The insured must show something more. Without evidence of a motive or “useful purpose” behind the risk-

\begin{flushright}
\textsuperscript{185} The question of onus facing the insured and the insurer is summarized by Osborne A.C.J.O., speaking for the Court on this point, in \textit{R.W. Hope Ltd. v. Dominion of Canada General Insurance Co.} (Ont. C.A.): “The onus is on the insured to establish that, on a possibility basis, the allegations made by the plaintiff, if proved, bring the claim within the four corners of the relevant policy. Once that threshold is met, the onus shifts to the insurer to show that the claim made falls outside the coverage provided by the policy because of an applicable exclusion clause. If there is an exception to an exclusion, the insured bears the burden of establishing that the exception applies.”


\textsuperscript{187} \textit{Supra} note 72.

\textsuperscript{188} \textit{Ibid} at para 6.
\end{flushright}
taking, the risk looks entirely inexplicable. Many people use methadone as a painkiller or to help stabilize the withdrawal symptoms of morphine or heroin or simply for pleasure. Its use, however, attends obvious risks. We do not know the purpose of those risks in this case. The judge rejected the evidence of the insured’s reputation as a drug user. If reliable evidence had been tendered of the insured’s history as a drug user, the result might have been different. Such evidence could have suggested that the insured had familiarity with the risk. It would have, moreover, provided an explanation for his risk-taking and taken his conduct outside of the category of superfluous or inexplicable risks.

**Conclusion**

In this chapter I have argued that the case law defining the term “accident” evidences a concern for the character and purpose of an insured’s risk-taking. Judges seem to employ two meanings of the phrase “courting the risk”. First, judges consider the insured’s *ex ante* perceptions of the likelihood of loss. Where a risk is seen as being highly likely to materialize, an insured can be deemed to have “courted” or “expected” subsequent losses. Second, where losses materialize as the result of purposeless risk taking, the resulting losses are, again, deemed “courted” or “expected”. This concern for the character or benefit of risk-taking does not fit with the *subjective expectations* test outlined in *Martin*. Like H&H's *abnormality*, the phrase “courting the risk” seems to operate as a shorthand reference for conduct judges want to deter. Superfluous risks, all other things being equal, are no more likely to materialize than gainful ones. The concern with the benefit of risks does not further an inquiry into *expectations*. 
A cost-benefit analysis underlies the putatively factual assessment. The analysis might be liberally applied. Coverage isn't denied to a party simply because the expected costs of a risk don't add up to its expected returns. On the other hand, an insured is impliedly required to demonstrate some expected return. It may be possible to formulate a test to make this doctrinally explicit. An insured might be required to explicitly demonstrate a “useful purpose”. This requirement would bring doctrine in line with how the cases seem to be resolved. Such a standard, however, doesn't simply test facts. An occurrence would be deemed to have been “caused” by an “accident” both as a matter of fact and of policy. The notion that proximate causation is ingested with cost-benefit considerations ill-fits with theory of “causal maximalists” who insist on a separation between the two.

There is nothing disingenuous about suggesting that those who stand out, either as abnormal actors or as those whose risk-taking seems to carry no redeemable purpose, are more likely to be found as “causes” as a matter of common sense. People likely make these types of judgements all the time. There is, moreover, certainty nothing ignoble about trying to ensure that those who take excessive or corrosive risks account to their victims or bear their own costs. There is a danger, on the other hand, of collapsing all of these types judgements into one, designating them as facts, and placing them beyond normative scrutiny.
Instrumental Causation, Efficiency and Insurance Law

The Law and Economics literature is less concerned by the danger of using causal concepts to further instrumental objectives. Guido Calibresi states this expressly: 189

If causal concepts can be used flexibly to identify the pressure points most amenable to our social goals, then use of such concepts has great advantages over explicit identification and separation of the goals.

In the Law and Economics literature causation is of a secondary or contingent importance to the social aims the law promotes. Law and Economics scholarship starts with the premise that legal rules and formulas only have value in so far as they advance social welfare. The doctrine of causation, accordingly, is only valuable in so far as it is able to do so. The idea of there being any inherent justice in holding responsible parties to account is discarded in favour of considering the distributional and efficiency implications of limiting liability to those causally responsible for a loss.

The literature, however, does not speak with a single voice on what “cause” actually means. Sometimes the doctrine of causation is tested against efficiency criteria. Sometimes causation is, in itself, something conceptually malleable enough to be shaped by instrumental objectives. The statement “X causes Y” is sometimes seen as neither efficient nor inefficient. Rather, causation may or may not be efficient depending on how it is formulated doctrinally and applied in any given case. Sometimes causation is, as in Realist literature, said to simply mask the judicial preference to assign fault one way or the other.

__________________________

Consider the following three examples from important early papers on the economics of legal causation. In the paper, “An Analysis of Causation and the Scope of Liability in the Law of Tort,” Shavell tests the efficiency of limiting liability with causation doctrine under both negligence and strict liability regimes.\(^\text{190}\) He does not, on the other hand, suggest that the concept simply bends to accommodate efficiency concerns. Calibresi, as suggested from the quote above, is more willing to do so. In the paper, “Concerning Cause and the Law of Tort”, he divides concepts of causation into “causal linkage”, “‘but for' causes”, and “proximate causes”.\(^\text{191}\) Calibresi suggests that the “flexibility” of all three allows them to serve our instrumental goals.\(^\text{192}\)

Causal linkage and but for case are functionally related to few, but fundamental tort aims. Proximate cause is, however, related to all such goals. It is little wonder then that it has always seemed the most complex, paradoxical, but also 'flexible' and policy-based of the causal requirements. Indeed, its very flexibility and explicitly functional policy orientation sometimes serves to hide the equally functional, rather than absolute, roles played by the other concepts.

In the paper, “Causation in Tort Law: An Economic Approach,” Landes and Posner, on the other hand, suggest the causation requirement is simply unnecessary.\(^\text{193}\) Tort law operates to deter parties from engaging in behaviour which increases the likelihood of

\(^{190}\) S. Shavell, “An Analysis of Causation and the Scope of Liability in the Law of Tort,” *Journal of Legal Studies* 9:463. He employs a doctrine of causation-in-fact that looks a lot like Wright’s. Whereas Wright asks whether a loss would have occurred but for the tortious aspect of the defendant's conduct, Shavell asks whether an accident would have occurred but for the defendant's failure to alter his or her level of care: at 467.

\(^{191}\) Calibresi, “Concerning Cause”, *supra* note 189.

\(^{192}\) *Ibid* at 80.

harm, rather than simply to deter parties who “cause” loss. A party whose conduct is a “but for” factor in an occurrence, but does not contribute to its likelihood, they argue, is not liable for the loss. They use the example of a trolley car that was struck by a falling tree. Suppose the operator was running a trolley at an excessive rate of speed at the time it was struck. Had it not been for the speed of the trolley, the accident would not have occurred. On the other hand, the speed made it no more or less likely that the trolley would be struck. They suggest it is futile to argue about whether the speed “caused” the accident. While such issues are determined under the rubrics of causation, courts, in essence, simply balance the value of the risk and the administrative costs against the cost of the injury.\textsuperscript{194} Causation, in Landes and Posner, is simply a (either conscious or unwitting) ruse.

It is, accordingly, dangerous to speak of the doctrine of causation as if it were a singular concept in Law and Economics. Whether used in any particular argument as a concept with an independent philosophical heritage, a “flexible” concept that is broad enough to promote instrumental concerns, or a mask for those concerns likely depends on the author or article. On the other hand, Law and Economics almost unwaveringly\textsuperscript{195} suggests the attribution of responsibility in law ought to be forward-looking and promote instrumental concerns. The doctrine of causation, whether a ruse or not, is subordinate to these aims.

This can occur under the rubrics of causation. But the instrumental concern

\textsuperscript{194} Ibid at 119.

\textsuperscript{195} For a dissent see, for example, Martha Nussbaum, “Moral Limits of CBA”, in Matthew Adler and Eric Posner (eds) \textit{Cost-Benefit Analysis: Legal, Economic and Philosophical Perspectives}, (Chicago: University of Chicago Press, 2001) at 193.
predominates.

The following chapter considers how Law and Economics scholarship resolves the question as to whether a loss is “caused” by a fortuitous accident within the scope of a third party liability insurance policy. As is readily evident, the question of causation is really entirely subsumed by the question of efficiency. The chapter does not argue against the subordination of causation doctrine to bald instrumentalism. It does, however, suggest that the concept of efficiency as an alternative to causation may be just as ambiguous and broad enough to support competing normative perspectives or interests. If the central concern of attributive judgements is the promotion of value, the following chapter questions how the courts ought to conceptualize value.

Law and Economics assumes people will rationally pursue their own value or utility. On the other hand, the empirical evidence available from an insurer's actuarial record suggests certain classes are more liable to losses than others. This suggests that such classes either have unique values which they pursue or that they act less-than-fully rational. The determination of whether a loss was “caused” by an accident or not is based upon a calculus, in orthodox Law and Economics, between the value and cost of the risk. This chapter questions whether this assessment ought to consider either variegated notions of the value people pursue or the less-than-rational way in which people pursue it.
Variety and the Law and Economics of Liability Insurance

Daydreams can be very consistent indeed.

- Amartya Sen

For an economic model to be more than a daydream it requires empirical validation. A model with empirical support might provide a rough sketch of how people behave in the real world. These sketches, however, are always rough. It is a dream to imagine that any economic model could accurately describe all of the nuances of human behaviour. There will always be statistical outliers that do not seem to behave, for whatever reason, like model actors. Empirical studies might suggest that a population’s behaviour imperfectly, but on average, conforms to an economic model. Economists do not need to respond to every particular instance that does not fit within a general theory. On the other hand, where economic models describe legal relationships, policy makers are left with the question of what to do with the outlier groups. We might first ask whether outliers are rational or irrational. They might be rational, though uniquely so. On the other hand, their ability to act rationally might be limited by irregular cognitive limitations or biases. If they are deemed rational, policy makers should ask whether the law needs to account for the ways in which they are uniquely rational. If not, policy makers should decide whether to intervene to correct irrational risk-taking. This chapter considers these three questions and outlines some of the normative concerns involved. It argues that the law accounts for the evident variety of legal actors in a diverse number of ways which are not simply subordinate to the principle of efficiency. Other principles accounting for variety, such as equity, proportionality, denunciation, or autonomy operate, often fundamentally, at odds with efficiency concerns.
Why Insurance?

Liability insurance offers a unique place to situate this discussion as it is readily theorized using both deductive economics and probabilistic actuarial theory. There is an evident tension between the underwriter's notion of risk as something determined by an individual’s age, gender, health, etc. and a legal economist's notion of risk-taking as something determined by economic incentives. This is not overly problematic in ordinary life. People are all unique and have distinctive tendencies, and yet we all make choices. Perhaps we make choices and pursue our interests within unique fields of opportunity, capacity, and frames of reference. This is not a particularly bold idea and does not seem to create any inherent contradictions. A tension, however, arises where this diversity is negotiated into law— or, at least, a law premised on rational choice theory. It raises the questions I have introduced. If law assumes rationality, ought it to accommodate or correct the actuarial outliers? I argue that economic theory cannot always provide an answer and that normative choices are needed.

Variety and Rational Risk-Taking

Some social activities involve inherent risks. Transportation, professional practice and industrial production, for example, will predictably generate accidents and losses. In these areas it is unrealistic that any regime could extinguish risk entirely. To some degree accidents and losses are accepted as unfortunate, but regular and inescapable. Economic theory suggests that people ought to participate in risky activities where the benefits
generated by these risks exceed the accompanying social costs of the accidents.  People are assumed to be aware of their exposure to fortuitous losses and are predicted to undertake an activity only if they derive a greater benefit from the activity than the expected costs they face. The costs generated by accidents, however, are not always exclusively borne by the person at fault. To promote efficient behaviour the law will assign liability so as to ensure that injurers internalize the entire costs of the harm they generate.

Under a negligence rule, legal economists suggest that the minimum standard of care ought to reflect the point at which the benefits of participating at a certain level of activity or degree of risk just barely exceed its costs. Where a defendant fails to meet this standard, he or she must satisfy the plaintiff’s claim. A risk-averse party, however, would rather pay for the costs of risk-taking in small, predictable premiums on insurance policies rather than run

---

196 The first full explication of this thesis in the context of tort law was Guido Calabresi’s *The Costs of Accidents: A Legal and Economic Analysis*, Yale University Press, 1970. For a brief review of the literature building from Calabresi’s model, see Steven Shavell, *Foundations of Economic Analysis of Law*, (London: Harvard University Press, 2004) at 192-193. This chapter uses Steven Shavell’s *Economic Analysis of Accident Law* (London: Harvard University Press, 1987) as a basic model and point of departure. While the key features of the model are discussed below, his book offers a significantly more detailed and nuanced account. Shavell’s model is used as it: (1) offers a clear and accessible example of an orthodox Law and Economics approach to the area of accident law; and (2) includes a discussion on the theoretical implications of the availability of insurance markets. A “conventional” or “orthodox” approach to Law and Economics is best understood with reference to the “Chicago School” of Law and Economics. Notable scholars include Richard Coase, Guido Calabresi, Henry Manne, Gary Becker, and Richard Posner. See Nicholas Mercuro and Steven Medema *Economics and the Law: From Posner to Post-Modernism and Beyond*, 2nd ed., Princeton: Princeton University Press, 2006 at 94. Mercuro and Medema suggest the literature is centered around the following assumptions: “(i) individuals are rational maximizers of their satisfactions in their nonmarket as well as their market behavior; (ii) individuals respond to price incentives in nonmarket as well as market behavior; and (iii) legal rules and legal outcomes can be assessed on the basis of efficiency properties” at 102.


198 Ibid at 6-9.

199 A party will be risk adverse if he or she experiences *marginally diminishing* enjoyment from each unit of wealth accumulated. For example, a party suffering marginally diminishing utility returns would be less affected by a loss of $100 if their net worth was $10,000 than if their net worth was $1,000. This party is risk adverse because he or she would rather, for example, spend $100 every month for insurance premiums than experience a $10,000 loss all at once, regardless if such a loss was expected to occur once every 100 months.
the risk of incurring a large, immediately payable judgement debt. If insurance was unavailable, such parties would take an excessive and inefficient amount of precautions so as to avoid incurring liabilities. Where insurance is available, on the other hand, parties are able to insure against this risk and participate in an efficient level of risk-taking. Legal economists suggest that in competitive markets insurers will charge a premium reflecting the expected costs of a risky activity.

There are additional economic considerations, however, if we make the rather obvious assumption that people manage their risks in different ways and that some people are more disposed to loss than others. The theory of adverse selection suggests that if risks are not differentiated, those less exposed to risk will disproportionately pay out the losses of those with greater risk exposures. Accordingly, “good risks” (or, as the phenomenon advances, “medium risks” and “not-quite-as-bad risks”) will progressively leave the insurance pool to avoid subsidizing “bad risks”. Eventually the pool will be so drained for money that the

200 Shavell argues that insurance generates value for society under two conditions: (1) if people are averse to bearing the risks of incurring awards of damages; and (2) if courts, at times, overestimate the level of precautionary behavior required in a situation. If courts are generally, but imperfectly, able to determine whether a party has taken optimal precautions, a party would be exposed to liability only in instances of judicial error. In the absence of insurance markets, risk adverse parties will take an excessive amount of precautions so as to avoid any appearance of negligence. This would be a suboptimal outcome. If insurers can underwrite the risk of judicial error, on the other hand, risk adverse actors can purchase insurance, spread the risks of bearing (incorrectly apportioned) legal liabilities, and, thus, participate at an optimal level of risk-taking. *Ibid* at 212-213.

201 *Ibid* at 186-191.


203 Adverse selection only operates in systems of voluntary, contractual insurance. Where insurance is, on the other hand, mandatory or free, all classes of risk will (or must) participate in mutual spreading. Such an arrangement socializes risk across the mutual pool, and is informed by principles of solidarity rather than contract. Richard V. Ericson, Aaron Doyle and Dean Barry, *Insurance as governance* (Toronto: University of Toronto Press, 2003) at 6.
principle of risk spreading will cease to operate and the market for insurance will collapse.\textsuperscript{204}

In order to distinguish the “good risks” from “bad risks”, insurers maintain considerable records of past insurance claims and accident histories that are organized into actuarial categories such as age, health, sex, occupation, etc. The data is used to rate individual risk exposures and ensure all members contribute premiums which approximate the actual amounts that they are predicted to draw from the pool.

The economics of insurance, thus, assume both the rationality of its members and their variety. Policyholders are both rational and unique. This poses a problem for economic models which assume actors to be homogenous. On the other hand, the variety in the way actors respond to risk can be reconciled with a theory of rational choice by suggesting that insureds are situated differently, hold different capabilities, or possess different understandings of the happiness they pursue. This variety can readily be built into economic models.

If we assume, for example, that an individual’s income corresponds to the time he or she spends at work, a person who earns a higher wage has a greater monetary incentive to speed. If women earn less than men, we might explain differences in claims histories based on the different monetary incentives to speed. In the real world, this is likely a glib or overly-

\textsuperscript{204} The nineteenth century, for example, saw the failure of a number of “Friendly Societies” that offered life insuring arrangements between guild or trade members but failed to differentiate between members on the basis of age. Their failure is often attributed to the problem of \textit{adverse selection}. See Bentley Gilbert, “The Decay of Nineteenth Century Provident Institutions and the Coming of Old Age Pensions in Great Britain,” 17(4) Econ. Hist. Review (Series 2) 550 (1965). Tom Baker, on the other hand, argues the collapse of the fraternal orders was due, rather, to shifting moral commitments: \textit{University of Connecticut School of Law and Working Papers}, Paper 3 (2001).
simplistic answer. However, for the sake of the argument, let us suppose that, excepting wage differentials, men and women are more or less the same. Because of the additional incentive to speed, men would, on average, optimally travel at a higher rate of speed and assume a higher degree of risk.

Suppose a man named Tom and a woman named Tammy both live in Chilliwack, BC and need to commute approximate 100 km from Chilliwack to Vancouver each work day. Tom is paid $200 an hour as a VP of a Vancouver-based mining company, while Tammy is paid $20 an hour to work at the company’s computer helpdesk. Tom and Tammy both devote 10 hours a day to working and commuting to work. During this 10 hour period, any time not spent commuting is spent working. If either Tom or Tammy drive an average of 80 km/h between highway and city driving, they will need to spend 2.5 hours a day commuting to work and back. If, on the other hand, Tom or Tammy average 60 km/h, they will need to spend just under 3 hours and 20 minutes on the road. Accordingly, by driving at the faster rate of speed the parties gain an extra 50 minutes a day that would otherwise be spent commuting. This is time that could be devoted to work. Given his hourly wage, Tom, therefore, has a $166.67 incentive to drive at the faster rate. Tammy’s incentive, on the other hand, is only $16.67. If the expected accident costs associated with driving at 80 km/h instead of 60 km/h was somewhere in between $166.67 and $16.67, say $100, we would expect Tom to bear the extra risks and Tammy to abstain from these risks. If parties were held to account for the expected accident costs produced from their faster rate of driving, Tom would still stand to gain an average of $66.67. Tammy, on the other hand, would expect to lose, on average, a net of $83.33 by driving at the faster rate.
The capacities for managing risky situations might also vary from person to person. Imagine, for example, that actuarial data suggested that teenage drivers are, on average, involved in more collisions than adult drivers. By factoring in disparities in the parties’ driving abilities or capacities into the expected costs they face for any particular level of risk-taking, we would, again, find that actors respond to incentives differently. The older and experienced driver will optimally travel at a faster rate than the more accident-prone, inexperienced youth.

Suppose that Tom has a gifted and successful 17 year-old daughter named Chrystal who also makes $200 an hour as a television actress. Chrystal still lives at home and also devotes 10 hours a day to transportation and work. As noted above, given their wages, both father and daughter have a $166.67 incentive to drive from Chilliwack to Vancouver and back at 80 km/h instead of 60 km/h. Suppose, however, that due to Chrystal’s inexperience she is 50% more likely to cause an accident than Tom. If Tom is, thus, expected to cause $25 of accident losses for every 5 kilometres he drives in excess of 60 km/h up until a maximum speed of 80 km/h, Chrystal, on the other hand is expected to cause $37.50 of accidental losses per every 5 kilometres in excess of 60 km/h. If both parties were held to account for their losses we would expect Tom to drive at 80 km/h. He would still gain $66.67 in doing so. Chrystal, on the other hand, would only drive at 75 km/h. At this speed she would expect to spend just under 40 less minutes on the road than she would if she drove at 60 km/h and, thus, be able to earn $133.33 more. At this rate of speed she expects to lose $112.50 more in expected accident costs. She should thus earn an average of $20.83 more money by speeding at
75km/h. If she raised her speed to 80km/h, however, she would be able to earn $166.67 more at work but would lose $150 in expected accident costs, thus only netting a gain of $16.67 from speeding. By the same calculation, if she dropped to 70km/h she’d only expect to gain $20.24 from speeding. Assuming Chrystal has a choice of driving at 60, 65, 70, 75 or 80 km/h, her optimal rate of speed is, thus, 75km/h.

We might also vary economic assumptions to take into account differences in the enjoyment people derive from the pleasure or adrenalin gained from speeding itself. Imagine that the actuarial record indicates that sports car drivers are more accident-prone than minivan drivers. We might assume that this is because sports car drivers enjoy speeding more than minivan drivers. We could include the thrill of speeding in the benefits the parties derive from driving to, again, help better explain differences in the claims-histories between the two groups.

Imagine that Benjamin and Baxter are partners at a Vancouver law firm and that they also make the commute each day between Chilliwack and Vancouver. They both make $100 an hour. Benjamin drives a Honda Odyssey (minivan). Baxter drives a BMW 650 (sports car). Benjamin is practical-minded and does not feel any particular enjoyment when speeding. If the expected accident costs are $25 for every extra 5 km/h of speed in excess of 60 km/h, Benjamin’s optimal rate of speed is 65 km/h. By driving at 65 km/h an hour he saves just over 15 minutes than if he travelled at 60km/h. This translates into a benefit of just over $25 an hour. As this benefit exceeds the $25 he expects to lose by travelling at 60 km/h, he will opt for the additional pace. A pace of 70 km/h would only translate into a savings of a few
minutes under half an hour and, thus, less than $50. It is not worth the $50 of expected accident losses to take the extra 5 km/h. Baxter, on the other hand, derives pleasure from speed. In fact, he enjoys speed so much that he is willing to pay incrementally more for every extra 5 km/h he drives in excess of 60 km/h up to a maximum of 80 km/h. This pleasure is worth $2 to him when he averages 65 km/h, $4 when averaging 70 km/h, $8 at 75 km/h, and $16 when averaging 80 km/h. If we factor Baxter’s love of speeding into the equation, we would expect him to travel at 70 km/h. At this pace he can shave just under half an hour from his commute compared to if he travelled at 60 km/h. This is worth about $47. He expects to incur $50 in expected accident losses by travelling at this rate. However, he derives a benefit of $4 from speeding. Accordingly, he receives a total benefit of $51 but only incurs a cost of $50. At 75 km/h, he would save $66.67 worth of his time, earn $8 dollars worth of speed benefits, but lose $75 in expected accident losses. His expected losses would outstrip his gains at this point. His optimal rate of travel is 70 km/h.

**Variety and Tort Standards**

These three examples suggest that rational actors with different external incentives, capacities, or inclinations will manage their risks differently. Where the accident costs of these parties are borne, at least in part, by victims, the law must disincentivize actors from taking excessive risks. This raises questions as to how the evident variety in optimal levels of risk-taking should affect or interact with the distribution of legal liabilities and the implementation of legal standards. In the conventional wisdom, speed limits are set so as to balance the time savings generated by traveling at particular rates of speed against the human
and property costs of collisions. They are, also, notably, uniform. These limits apply across a population, regardless of income-level, experience, driving history or disposition. They do not consider the individual utility of different classes of drivers. This, however, has efficiency implications. Where a uniform standard applies, the parties deriving a benefit in excess of the standard would run the risk of being found negligent and, if they were risk-averse, would take an inefficient excess amount of precautions to avoid such a finding.

If judges were to set legal standards so as to maximize the total social benefits of an activity, those persons exacting the greater benefit would be afforded a lower standard of care. The law would be individualized. As the examples above indicate, the person who forgoes the higher wages from time spent on the roads or the person who derives excessive utility from speed enjoys a higher benefit from driving faster. Accordingly, their expected accident losses would balance against disproportionately higher benefits of speed and would, thus, permit a lower-than-average standard of care.

This is not how the law generally works. One reason for this is that it is often prohibitively difficult to quantify the benefit that people derive from risk-taking. Landes and Posner, for example, describe the burden of assessing the character of the benefit an actor receives

205 Charles Lave, on the other hand, influentially argued that average rates of speed have no statistical bearing on fatalities. Rather, it is deviance from the speed of traffic that generates accidents. Accordingly, he suggested enforcement should focus on deviance rather than rates of speed. “Speeding, Coordination, and the 55 MPH Limit” The American Economic Review, 75:5 (1985) 1159. This chapter, for simplicity, assumes that the conventional wisdom holds.

206 Shavell argues that a strict liability rule does not need to balance the costs of an activity against its benefits. However, unless this rule is supplemented with a defence of contributory negligence, victims have no incentive to take precautions from their end. If a strict liability rule is supplemented with a defence of contributory negligence, however, courts must still assess negligence and, thus, determine the costs and benefits inhering to an activity. Strict liability rules are not considered in this chapter. supra note 196, at 24-25.
from an activity as an information cost.\textsuperscript{207} Information costs require judges to limit the subset of factors they take under consideration when determining standards and their applicability to the facts of any given case.\textsuperscript{208}

\textit{Information costs} are not, however, the only reason why uniform standards are imposed. It would strike most of us as an affront to principles of equity and fairness if the law let the rich off in situations where it held the poor to account. It would seem a perversion in principle to give a clemency to people deriving a pleasure from the very thing the law seeks to restrain. A uniform, objective standard, on the other hand, is, at least ostensibly or aspirationally blind to privilege or disposition. It holds all members of society to account for the harm produced by risk-taking. This uniformity has a moral and not merely economic import.\textsuperscript{209}

\begin{flushright}
\footnotesize
\textsuperscript{207} Landes, William and Posner, Richard, “The Positive Economic Theory of Tort Law” \textit{Georgia Law Review}, (1981) 14: 851 at 876. As will be discussed below, insurance purchasing decisions might very well indicate individually-specific benefits for, as an example, road travel, and make such costs negligible.

\textsuperscript{208} Shavell, for example, suggests courts assess the costs of readily quantifiable precautionary behavior, but not the hard-to-measure costs of foregoing higher levels of activity. Actors, accordingly, have an incentive to undertake optimal precautionary measures, but will engage in excessive levels of activity. Shavell, supra note 196 at 24-25. Landes and Posner draw a similar conclusion. They write: “One way of economizing on the costs of determining (the standard of care) is for the court to look at only a subset of care inputs—those relate to the safety with which the activity is carried out, as distinct from the amount of the activity.” Landes, William and Posner, Richard, \textit{Ibid} at 21. The distinction between precautionary behavior and level of activity or safety and amount, however, is not intuitively clear. Does, for example, the speed of one’s driving signify participation at a certain level of activity or a certain level of precautionary behavior? On one hand, the speed a person travels when s/he does travel is clearly distinct from the number of times a week a person travels. In this sense, speed looks more like a “precaution” than an “amount” and, thus, within the competence of courts to determine. On the other hand, an assessment of the costs incurred or, conversely, the value generated by speeding would require the court to consider the “character of the benefit” of the time-value, or intrinsic benefit produced by speeding—an “amount” type of assessment. The distinction between precautionary measures and levels of activity is noted but not specifically addressed here.

\textsuperscript{209} The adoption of a uniform standard around the utility preferences of an average member of the insurance pool implies governance through norms. For a more equivocal reading of the moral import of governance through norms in insurance law see Francois Ewald, “Norms, Discipline, and the Law” \textit{Representations} 30 (Spring 1990) 138.
\end{flushright}
Variety and Insuring the ‘Monstrous’

The evident variety in the ways people handle risk is likely influenced, in part, by the different ways people take pleasure from risky pursuits. If people do take their pleasure differently, the law needs to ask how disparate forms of pleasure are commensurate. In our third hypothetical model, sports car drivers enjoy a greater utility from speeding than minivan drivers. They are led to take an excessive amount of risk in pursuit of this pleasure. If all types of pleasure rank equally, the courts ought to afford sports car drivers a greater liberality on the roads than others. It seems unlikely a judge would do so.

Even if the information costs of determining the sport car driver's utility from speed were insignificant or surmountable we still might say that an injurer's “speed utility” should never measure against the disutility of being victimized in a traumatic automobile accident. Courts might be justified in denying or limiting access to insurance agreements designed to facilitate this lesser form of pleasure at the expense of a victim's personal integrity. A court could, for example, exclude this pleasure from calculus of costs and benefits in consideration of the optimal standard of care under a negligence rule or of the extent of coverage purchased by an insured in a broadly-worded “accident” policy.210

210 Most policies that are widely held by the public are contracts of adhesion drafted in standard language and offered on a take-it-or-leave-it basis. In many circumstances insurers issue general liability or all-risk policies affording policy holders broadly drafted coverage for liability incurred by way of, for example, “accident”. Where an insured is found negligent, a court is, thus, required to determine whether this negligence was “accidental” per the insuring agreement. A ‘rational’ insured would purchase an amount of insurance allowing him/her to participate at a level of risk-taking where the expected returns from an activity exceed its expected losses. If individuals arerationally unique, the expected returns and, thus, the amount of insurance that each uniquely situated insured ought to hold will be different. The coverage decision, thus, requires courts to assess the subjective intentions or expectations of the contracting parties. Scott v. Wawanesa Insurance Co. [1989] 1 S.C.R. 1445 at para 12 (S.C.C.). In order for a court to optimally construct an insurance policy and allow an insured to participate in his or her individually-optimal level of risk-taking it must have an understanding of the private benefits expected by an insured from the risk-taking.
Contingent and Hedonic Valuation

The question of the rank of pleasure has a long heritage in utilitarian theory. In 1781, Jeremy Bentham argued that pleasure ought solely to be considered as an empirical matter. Pleasure is measured only with regard to its intensity, duration, certainty or uncertainty, propinquity or remoteness, fecundity, purity, and, extent. J.S. Mill, on the other hand, distinguished pleasures of the higher faculties from those of the body. A court, following Mill, would be bound to consider, not only the quantity, but the quality of the activity under question when determining a tort standard or construing an insurance policy, for example. If there is an answer to Bentham and Mill’s question it continues to elude economists today. Economics has a number of empirical tools to value and commensurate pleasure without resort to Mill’s teleology between mind and flesh or some other qualitative ranking. Contingent valuation methods, for example, determine value by surveying stakeholders on the money-value they would hypothetically pay for projects or proposals. This is an imperfect tool. Survey

211 Ibid at 30-31.
214 Eric Posner, for example writes: “One could imagine that welfare is objective in the sense that regardless of one’s preferences one has less welfare if one unknowingly takes high risks tan if one does not. Or welfare could involve not just having any preferences but the right kind of preferences, and while a person’s preference for driving is not sufficiently informed, that preference does not fully count in the person’s welfare. But these are difficult and complex problems that have not been resolved by economists and philosophers.” “Probability Errors: Some Positive and Normative Implications for Tort and Contract Law” in Francesco Parisi and Vernon Smith (eds) The Law and Economics of Irrational Behavior (Stanford: Stanford University Press, 2005) at 470.
respondents seem to disproportionately weigh losses over gains,\textsuperscript{216} discount small probabilities, or place a similar value on all large, though significantly discrepant, consequences,\textsuperscript{217} for example. Due to the artificial way in which surveys construct value, the contingent valuation method is often used only where economists are unable to trace consumer choice through market behaviour.\textsuperscript{218}

\textit{Hedonic} methods, on the other hand, rely upon traceable behaviour. Hedonic valuation infers the value of a benefit from consumer decisions and observable market interactions.\textsuperscript{219} If consumers are willing to pay \(x\) number of dollars to live in quiet neighbourhood rather than a noisy neighbourhood and all other factors between the two neighbourhoods are equal, we can infer that \(x\) represents the value of noise reduction, for example.\textsuperscript{220} Money is used as an absolute metric of utility.

\textbf{Wealth Maximization}

Richard Posner offers a paradigmatic theorization of the \textit{hedonic} approach.\textsuperscript{221} He argues that

\begin{quote}
\textsuperscript{218} Portney, \textit{supra}, note 215 at 14.
\textsuperscript{219} The “revealed preference” method for determining preferences was pioneered by Paul Samuelson, “A Note of the Pure Theory of Consumer’s Behavior” \textit{Economica}, 5 (1938). Samuelson writes: “The individual guinea-pig, by his market behavior, reveals his preference pattern—if there is such a consistent pattern” quoted by Amartya Sen “Behavior and the Concept of Preference” \textit{Economica} 40:159 (1973) at 241. The \textit{hedonic} determination of implicit preferences from explicit purchasing decisions was developed, for example, by Sherwin Rosen, “Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition” \textit{Journal of Political Economy}, 82:1 (Jan – Feb 1974) 34 – 55.
\textsuperscript{220} Ibid.
\textsuperscript{221} Richard Posner, “Utilitarianism, Economics, and Legal Theory,” 8 Journal of Legal Studies (1979) at 103-105
\end{quote}
the classical utilitarian theory contains two “monstrosities” largely mitigated if money is considered to be the sole measure of happiness. First, he suggests that without some universal metric, it is impossible to commensurate qualitatively different pleasures. Second, utilitarianism may “sacrifice the innocent individual on the “alter of social need”. Utilitarianism might allow the retributive use of torture, for example, if public vengeance was sufficient to outweigh the pain of the tortured. Posner suggests that we ought only to quantify costs and benefits with reference to what people have actually paid for social goods or would be hypothetically willing to pay in situations where markets do not exist. We can refuse to recognize the utility of individuals who appropriate value for themselves without paying for it in return and, thus, do not need to consider the “monstrous” utility of the thief, rapist, or murderer. Posner suggests, moreover, wealth maximization is not contrary to liberal notions of person inviolability. As individuals have both personal and productive interests in their bodies, while a slave owner’s interest is strictly productive, individuals would be willing to pay more for the rights to their bodies than slave owners. A wealth maximizing approach would, thus, afford them this basic right.

There are obvious class implications in Posner’s wealth maximization. If money is a universal signifier of value, the interests of the well-situated outweigh, a priori, the interests of the dispossessed. To use our first example, a law informed by wealth maximization

\[222 \text{ Ibid.} \]
\[223 \text{ Ibid.} \]
\[224 \text{ Ibid at 125.} \]
\[225 \text{ Ibid.} \]
\[226 \text{ Jeanne Schroeder, for example, attacks Posner’s conclusion that wealth maximization would necessarily allow women to repurchase the right to their bodies to avoid rape in Richard Posner, Economics of} \]
would hold low income drivers to the more onerous and time-consuming standard of care because of their lower pecuniary interest in time-savings and would, thus, ask the poor to account in circumstances where it pardons the rich. This, again, clearly offends principles of equity hardly peripheral to the law.

However, there is a more fundamental problem that makes empirical valuation techniques ill-fit to identify and explain the values relied upon when judges weigh the character of insured activities. Suppose the case law suggests that judges are unlikely to value risks which seem to have been taken gratuitously and hold no redeemable value excepting the sheer thrill of “courting the risk” itself.\textsuperscript{227} Suppose, for example, that judges refuse to recognize the utility enjoyed by those speeding and apply a uniform tort standard to both sports car drivers and minivan drivers. In the absence of an insurance market it might be possible to assert that the value of the thrill of speed is too hypothetical or speculative to warrant consideration in the courts. However, where an insured pays a differentiated premium covering the extra liabilities expected from speed, the value of speeding, under Posnernian theory, is clear. It represents the amount of the additional premium. There are no information costs incurred. The price an insured is willing to pay to enjoy their risk-taking is wholly reflected in the amount paid on the policy. A wealth maximizing court would, thus, have little excuse for not

\textit{Justice} (Cambridge: Harvard University Press, 1981) at 71. She argues that the value a man would be willing to pay for access to the body of a non-consenting female could, in theory, outweigh what the female would be able to produce to protect her sexual integrity: “It is not enough that the woman value her sexual autonomy—she must be able to pay for it. In contradistinction to Posner’s analysis of the case of slavery, it is hard to argue that the woman would be able to borrow the purchase price, since the value of her bodily integrity to her is not related to productive capacity per se.” Jeanne Lorraine Schroeder, \textit{The Triumph of Venus: The Erotics of the Market} (University of California Press, 2004) at 236. See also, Ronald Dworkin, “Is Wealth a Value?” \textit{Journal of Legal Studies}, 9:2 (1980).

\textsuperscript{227} See Candler (1963), 40 D.L.R. (2d) 408.
considering the value inhering to that individual when speeding.

To take another example, suppose a National Hockey League player desired comprehensive coverage for any contingency associated with the game. Suppose that, during the heat of the moment, hockey players, at times, transgress the customary limits on the aggression deemed implicitly part of hockey and participate in criminal violence.\(^\text{228}\) If a player is risk-averse and the value to the player and to spectators of this highly aggressive style of play outweighs expected injury costs, it is efficient if he has insurance coverage for the unintended consequences of his potentially criminal conduct on the ice. Imagine the player avenges himself on an opposing player by striking from behind. The attacker strikes with the intent to “hurt”, but not to “injure”. When the victim falls, his head strikes the ice and he suffers a serious concussion and vertebral ligament damage. This is criminal.\(^\text{229}\) The consequent losses were fortuitous, unintended and, thus, might have been underwritten. They are, however, not recoverable by virtue of the \textit{criminal forfeiture} rule. Under this rule, an insured cannot be indemnified for intentionally-caused losses, committed in contravention of the criminal law.\(^\text{230}\) This rule is not dependent upon the rights or obligations set out in the contract itself, but, rather, is imposed, in the words of Lord Denning, because “it would shock the public conscience if a man could use the courts to enforce a money claim either under a


\(^{229}\) R. v. Bertuzzi 2004 BCPC 472.

\(^{230}\) Sirois v. Saindon [1976] 1 S.C.R. 735. The rule is codified in provincial insurance statutes. See, for example, British Columbia’s \textit{Insurance Act} [RSBC 1996] C. 226, s. 28. Notably, this limits the common law \textit{criminal forfeiture rule}. Insurers are unable to deny recovery on the grounds that loss occurred as the result of an insured’s criminal act unless the insured intended the loss.
contract or a will by reason of his having committed such act.”231

A wealth maximizing court, on the other hand, would render the opposite result. Where the aggressive hockey player and his insurer agreed on a premium to cover liabilities for unintended consequences resulting from criminal hockey violence, the value of that extra degree of violence to the player is again, not speculative. It is the premium paid to participate in the risk. Assuming that a tort award fully compensates the victimized hockey player, a court would have to assume that the value generated by the insurance contract in facilitating such violence exceeded its costs.

Public policy limits such as the criminal forfeiture rule are, moreover, fundamentally at odds with wealth maximization. If, again, tort awards adequately restore plaintiffs, wealth maximizing courts would never apply public policy rules independently of the limits on risk-taking set out by the contract itself. The evident value of the contract to the parties in the premium would clearly outweigh the “speculative” value of proscribing insurance for amorphous reasons of public policy.

As insurance premiums seem to provide their own hedonic value, there is little economic (or Posnerian) explanation for why the law circumscribes the market in one way or the other or why it denounces and constricts activities with clear and discernible value. Perhaps, our assumption that tort awards provide full restoration is incorrect and that this provides part of the explanation as to why courts refuse contracts that facilitate dangerous activities benefiting

a limited class. The better answer, however, is likely that the law employs both market and non-market metrics of value. Perhaps, the courts rely implicitly on teleologies between high and low pleasures to assign value. Perhaps they omit “preferences based on ignorance and haste, preferences deformed by malice, envy, resentment, or fear”. 232 Perhaps, judges hold certain rights to be inviolable and, thus, not “valued” against social need.233 Or rank some values with an urgency and weight independent of their instrumental value.234 Perhaps, the judicial “weighing” of value is not always necessarily concerned with social consequences or maximizing total social utility.235 Or that judges realize that their decisions, in themselves, have a capacity to shape what and how things are valued.236 At any rate, the notion that judges exclusively balance monetized values with the singular concern to expand wealth is clearly wrong.

It is true, of course, that legal economists do not pretend to be able to reconcile all the nuances of judicial thought with economic doctrine. Rather, they are concerned with whether the outcomes rendered generally conform to economic theses.237 Nevertheless, the failure of an economics (based on market valuations) to account for clear judicial circumscriptions on

---


233 See, for example, Robert Nozick, State, Anarchy, and Utopia, (Perseus Books Group, 1977) chapter 3.


235 Nussbaum, supra note 232.


237 Richard Posner writes: “Granted few judicial opinions contain explicit references to economic concepts. But often the true grounds of legal decision are concealed rather than illuminated by the characteristic rhetoric of opinions.” Richard Posner, supra note 15 at 25.
“wealth”-generating contractual arrangements, such as the criminal forfeiture rule, puts law and economics’ descriptive thesis into real doubt. The cases suggest, both in rhetoric and outcome, that value is something clearly more than just what one is willing to pay. Value, rather, is an inherently normative concept.

Variety and Irrational Risk-Taking

Bounded Rationality

There may be another explanation for the evident differences in the ways or degrees that people take risks. Some people may simply be more prone to behave irrationally than others. Some people, for example, might suffer from cognitive biases which limit their foresight or temporally distort their desires. These biases might cause them to act, in spite of themselves, contrary to their own interests. Accordingly, they take excessive risks, are involved in more accidents and are charged higher premiums than others. This hypothesis would be inexplicable in traditional rational choice theory. In traditional theory, the assumption of rationality preconditions model-building. The behavioural economists, on the other hand, temper the assumption. They suggest rational behaviour is evident even within ostensible instances of irrationality.238 However, this rationality is constrained or “bounded”239 by the


239 The term “bounded rationality” was coined by Herbert Simon. He writes: “The term ‘bounded rationality’ is used to designate rational choice that takes into account the cognitive limitations of the decision-maker – limitations of both knowledge and computational capacity” in Eatwell, et a. (eds) The New Palgrave: Utility and Probability, (W.W Norton & Company, 1990) at 15. While Simon uses the term to describe how humans “satisfice”—or make decisions on necessarily incomplete information—the term is also used in the literature to describe systemic defects inhibiting agents from making fully rational decisions. “Maps of Bounded Rationality: Psychology for Behavioral Economics”, The American Economic Review 93:5 (December 2003) at 1449. The term is used in this chapter in the latter sense. Gerd Gigerenzer, however, discusses the difference between Simon’s concept of “bounded rationality” and the idea of “bounded rationality” as a cognitive limitation and suggests the latter distorts its original meaning. “Is the Mind Irrational or Ecologically Rational”
types of cognitive defects or limitations which have been theorized and tested by psychologists. The literature suggests that people can be prone to, for example, disproportionately favour outcomes which pay earlier dividends, over-rely upon stereotypes or past experiences when judging similar, though distinct, events, or weigh gains less than proportionate losses. A theme running through the literature is that boundedly rational actors favour outcomes that appear optimal in the short-term but, in fact, compromise their long-term interests. The challenge for a law and economics of bounded rationality is, thus, to structure incentives so that the boundedly rational actor subjectively perceives costs mirroring the actual objective costs of risk-taking.


240 “Hyperbolic discounting” or “myopia”. See, for example, George Ainslie “Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control” Psychological Bulletin, 82: 4 (July 1975).


243 “Risk Aversion”. Thaler, as noted argues that risk aversion makes hedonic pricing methods difficult. “Toward a Positive Theory of Consumer Choice” supra note 216. On the other hand, risk aversion, as Shavell argues, makes insurance economically viable. Is risk aversion a cognitive bias or something reconcilable with traditional rational choice theory? A party will be risk adverse if s/he experiences marginally diminishing enjoyment from each unit of wealth accumulated. There is nothing, a priori, however, to suggest why a party would have marginally diminishing returns over neutral or marginally increasing returns. For the very poor, however, marginally diminishing returns possess a biological urgency. The utility gained at the first moment of wealth is survival. The poor, however, are not involved with private insurance. They would hardly be expected to invest their money in the payment of insurance premiums instead of meeting their immediate survival needs. Here, the theory of marginally diminishing returns implies no relationship with risk. The poor must always bear their own risks. Imagine, on the other hand, an individual with a financial situation straddling deprivation and surplus. Every dollar added to this position merely adds to the state of surplus and, if spent, luxuriates the present. On the other hand, every dollar lost involves a deprivation and a threat to physical security and dignity. Thus, the relationship between marginally diminishing returns and risk aversion only develops with surplus. It also has an upper limit. Where an individual holds a sufficiently diversified investment portfolio and maintains assets that are large in relation to his/her risk exposures, expected losses can be borne in their entirety without having much impact on overall wealth. Accordingly, for the very rich, risk takes its true empirical form and can be managed through probability all but free from the threat of deprivation. Accordingly, marginally diminishing returns and risk aversion seems only to represent a cognitive bias for the very poor and the very rich.

244 Christine Jolls, Cass R. Sunstein, and Richard Thaler “Behavioral Approach to Law and Economics”
Consider the implications for our speeding example. Some studies have suggested that drivers tend towards higher-than-optimal speeds because they “hyperbolically discount” future, remote and probabilistic consequences against present returns.\(^{245}\) After driving at a high rate of speed, drivers become desensitized to the feeling and imagine that they are driving slower than they actually are.\(^{246}\) Speed works like an addictive good. The more speed you consume, the more you want. As past consumption of speed affects present preferences, speeders take more-than-optimal risks to satisfy what they experience as an optimal level of risk-taking. Suppose that this addictive or desensitizing effect is shared by the population at large. A number of strategies might be used to mitigate or dampen the effect. Fines could be doubled for speeding when coming off freeway exits onto residential streets, for example. Enforcement authorities could selectively target spots where speed limits drop and could advertise their intentions. The city could rely upon increased signage. When such strategies are applied to the population at large, they do not seem particularly problematic.

Suppose, on the other hand, that men suffer from a myopia which makes them less aware of the consequences of desensitization and that this myopia has a compounding effect on the addictive character of speed.\(^{247}\) Suppose, that the disproportionate number of claims made


\(^{246}\) The authors provide this illustration: “This shows that when you have been driving at a speed of 70 miles/h for nearly 20 minutes, 50 miles/h feels like 40 miles/h. When you have been driving at 70 miles/h for 35-40 minutes, 53 miles/h feels like 40 miles/h. In short, speed is almost addictive: The more you consume of it, the more difficult it is to reduce consumption to a targeted, lower level” *Ibid* at 91.

\(^{247}\) In the seminal article, “A Theory of Rational Addiction” Gary Becker and Kevin Murphy suggest that people suffering from hyperbolic discounting or myopic utility preferences are more prone to become hooked
by men against their auto-insurers is due to the operation of this cognitive defect.

Accordingly, enforcement, regulatory or judicial strategies to align subjective perceptions
with objective consequences of risk should differ between men and women. If strategies
solely target the non-myopic population who suffer desensitization to speed but are otherwise
rational risk-perceivers, they will under-incentivize the myopic driver to take appropriate
care.

Optimal regulatory strategies require segregating the population into myopic and non-myopic
risk perceivers and tailoring deterrents specifically to each group. The myopic group
requires special attention. In order for risk-taking to be optimized, the “addiction” and
“hyperbolic-discounting” effects on the irrational perceiver must be outweighed by increased
expected costs. They might be selectively targeted by increased pre-accident regulatory
strategies or through increased post-accident punitive awards. Either way, some mechanism
is needed to bring the true cost of risk-taking to bear presently on the mind of the myopic
actor.

Variegated Rationality

Variety, however, presents unique problems to a law premised on the economics of bounded
rationality. First, the information costs of ascertaining benefits and segregating litigants are
great. If people are presumed to be rational (and if the law were, in fact, committed to wealth
maximization) risk preferences could be gleaned relatively painlessly from people’s insurance

on addictive goods than others since they fail to fully internalize the future consequences of both the expected
losses from current consumption and the impact of current on future consumption. Addiction and myopia have
a compounding effect. Gary Becker and Kevin Murphy, “A Theory of Rational Addiction” The Journal of
premiums. The law could, theoretically, be diversified to reflect these preferences. Tailoring the law to account for various degrees of rationality, on the other hand, requires access to something more than just observable market behaviour. Where people are presumed to be boundedly rational there is no necessary connection between market decisions and their true preferences.

Behavioural economists suggest, rather, that people often make consumer decisions that conflict with their true interests. Where an individual, for example, purchases a level of insurance to indemnify against risks under the influence of “hyperbolic discounting”, we could not say that the premium paid reflects the actual value that he or she places on an activity. Accordingly, behavioural economists qualify or substitute hedonic values or “revealed preferences” ascertained from surveys and psychological testing. These methods, as noted, are inherently less certain than hedonic or revealed preference methods and are always subject to all the criticisms that can be levied against a survey’s experimental design or controls.

More fundamentally, Dan Kahan has recently argued that the data which could lead researchers to conclude that an individual suffers from a cognitive bias might also, in many instances, actually indicate an expression of normative commitments. He argues that people’s unwillingness to adjust perceptions and behaviour to reflect “objective” positions on

risk might suggest that people gain (hardly irrational) returns from the opportunity to take a position on their fundamental principles and beliefs. He offers these examples:

The individualist, for example, who continues to worry more about being rendered defenseless than about being shot as the risks of insufficient gun control appear to increase might “not so much [be] afraid of dying as afraid of death without honor.” Similarly, for the person who values an activity—say, smoking—precisely because she subscribes to an ethic that prizes the “authenticity of impulse and risk,” a cultivated disposition to discount the likelihood of personal harm may be integral to the very form of life that activity helps her to experience.

Kahan’s argument, thus, cautions against using corrected preference in the place of revealed preference metrics. While we still might wish to restrict access to guns or cigarettes, this requires more than simply dismissing the claims of the individualist or the free spirit as being irrational or inefficient. We would need to confront their normative positions with our own.

Moreover, if survey values revealing corrected preferences are used by courts, further complications arise. A judge must be certain that the controls operating at the time the values were rendered also exist in the court. This would likely prove to be too much. Take our example of male myopia and speeding. Suppose that a myopic male defendant faces the courts. The court might presume that the defendant was unduly affected by his failure to perceive the consequences of his actions. Specific deterrence might warrant that he bear a punitive award so that his perceived and actual costs mirror each other. On the other hand, suppose that the insured has an extremely lucrative position and, thus, a heightened incentive to drive fast. An optimal legal standard would reflect these incentives.

\[250\] Ibid at 114.
Notably, the two effects militate towards opposite conclusions. The “myopia effect”
demands a higher standard; the “income effect”, a more permissive one. Which of the two
effects explains his speed (and why he paid the premium he did)? The price of the premium
on risk represents both (rational) preferences and incentives and (irrational) risk perceptions.
Unless the values rendered in the survey were controlled specifically to account for income-
related incentives, the question as to whether his risk-taking and consumer behaviour reflect
an “income effect” or a “myopia effect” is impossible to say. It is hard to imagine many
cases where legal decision-making could furnish such controls.251

*Proscribing Irrationality*

Second, the prescriptions of a law and economics of bounded rationality seem problematic.
Individuals suffering from cognitive defects require an extra disincentive to bring the true

251 I am mindful, here, that two of the arguments in this chapter seem to militate towards opposite
conclusions. In the section, Variety and Rational Risk-Taking, this chapter argues that the information costs
associated with assessing the private benefit inhering to insured's can be overcome if we give hedonic value to
the price of the premium that an insured pays for the right to participate at a given level of risk. Given that an
insured is willing to pay this price on his or her risk, instances disallowing third party liability contracts are
irreconcilable with rational choice theory. In this section, on the other hand, I note that hedonic methods can
not be used to determine the preferences of boundedly rational actors and that that survey or contingent
valuation methods are unlikely to garnish sufficient controls to be suitable for the court room. This argument
suggests that the information costs of risk-taking are too steep and, perhaps, that courts ought not, as Landes
and Posner argue, consider the benefits accruing to parties, but rather, only the readily ascertainable costs of
precautions. On the one hand, these arguments seem to conflict. If actors are perfectly rational, information costs
are negligible; if they are not, such costs are prohibitive. The fact that the information costs are
considerable, on the other hand, does not allow for an efficiency explanation of why courts would interfere with
insurance contracts for reasons of public policy, for example, where the criminal forfeiture rule is applied. If
courts are unable to assess whether an insured desires coverage for criminal conduct due to irrational and
correctable preferences or due to uniquely rational preferences that ought to be taken into account by the law,
there is no justification for either prohibiting or allowing such contracts one way or the other. If a court allows a
contract, there is a possibility that the insured was an irrational or boundedly rational actor compelled to enter
into such a contract on the basis of a skewed self-interest. On the other hand, if a court prohibits such a
contract, there is a possibility that the insured was simply rational, though uniquely so, and that the prohibition
denies the insured the hedonic value in the risk that he or she contracted for by paying the insurance premium.
The answer, without furnishing considerable controls, is economically indeterminate. The law, on the other
hand, is not paralyzed by this uncertainty. It prohibits such contracts. This reflects denunciatory and corrective
purposes of law rather than simply its efficiency commitments.
consequences of risk-taking to bear on their decision-making. If an individual hyperbolically-discounts future consequences, an anticipated loss will have to carry an added weight to make its risk appear like it would for any other party. Accordingly, a mental weakness would seem to aggravate the circumstances of loss or, at least, demand the weightier reply. Optimal laws might, thus, increase sanctions in proportion to the limits of a defendant’s mind.252

The law, in fact, seems to do just the opposite. It is rare that penalties increase as the awareness of an actor diminishes. In ancient Athens, for example, the legislator Pittacus established that penalties should be more severe if committed under the influence of wine.253 There are few such modern examples. The regulatory state, of course, extensively guards against irrational participation in all sorts of areas. Those who lack the demonstrated capacity to appreciate and negotiate common risks are denied licensed access to risk-taking. Where the law must deal ex post with loss, a lack of capacity, on the other hand, generally mitigates the severity of sanctions.

252 Nathan Berg and Gerd Gigerenzer argue that governments do not necessarily need to respond to the boundedly rational actor with increased severity or paternalism. They suggest, for example, that, assuming sexual consumption is not something consumed addictively, a less restrictive stance on sexuality might actually reduce myopic risk-taking if people choose to meet their “sexual aspirations” with safer and more readily available substitutes such as pornography. They suggest that whether instances of bounded rationality warrant increased severity or restrictions depends entirely on contextual factors. Nevertheless generalization which suggests that bounded rationality implies increased regulation is overstated. However, where the behavioral economics does suggest that the law respond with greater severity or a heavier standard to those with a reduced capacity, it seems to talk at cross-purposes with important principles in law. Nathan Berg and Gerd Gigerenzer, “Psychology implies paternalism? Bounded rationality may reduce the rationale to regulate risk-taking” Social Choice Welfare 28:337-359 (2007) at 356-357.

In contrast to Pitticus, in our law, a crime committed while voluntarily intoxicated often mitigates a crime of general intent\(^{254}\) and excuses crimes of specific intent.\(^{255}\) In tort law, adults are held to a higher standard than children\(^{256}\) and professionals than lay people.\(^{257}\) Examples abound. The notion of proportionality or of the commensurability of sanctions with moral culpability exercises a tremendous, if not principal, influence in many instances.\(^{258}\) Proportionality would stand on its heads if a lack of capacity increased the severity rather than clemency of the law. Given that our law is concerned not solely with deterrence, but also with the apportionment of moral fault, a law and economics of bounded rationality ought to be cautious about its prescriptions in highly coercive fields.\(^{259}\) The use of increased sanctions to provide appropriate deterrence to the irrational risk-perceiver is, in many respects, foundationally at odds with how the law apportions moral fault.

**Paternalism**

Finally, the idea that the state should correct individuals who are unable to properly assess and act in their own interests is inherently paternalistic. It authorizes an increased state


\(^{255}\) See for example, *R v. Walker* 2008 SCC 34 at 18.


\(^{258}\) See, for example, Ernest Weinrib, “Towards a Moral Theory of Negligence Law” *Law and Philosophy*, 2, 37.

\(^{259}\) Colin Camerer, Samuel Issacharoff, George Loewenstein, Ted O'Donoghue, Matthew Rabin, for example, suggest that recommendations from behavioral economics ought to be applied to law asymmetrically. In situations where actors reveal genuine remorse for their decisions, governments might safely intervene. Governments, on the other hand, should show restraint where people are demonstrably more equivocal after ostensibly “irrational” decisions. Colin Camerer, Samuel Issacharoff, George Loewenstein, Ted O'Donoghue, Matthew Rabin, “Regulation for Conservatives: Behavioral Economics and the Case for ‘Asymmetric Paternalism’” *University of Pennsylvania Law Review*, 151:3 (Jan. 2003) 1211-1254.
presence and supervision over consumer decisions in a way at odds with traditional liberal theory. Welfare economics in general is, of course, itself at odds with the concept of a “night watchman state”. A libertarian would hardly suggest the state should be involved in weighing the interests of citizens against each other to promote overall happiness.  

Behavioural law and economics, on the other hand, suggests that even “interests” are subject to state inspection and reformulation. As this confers greater governmental involvement it ought to be met with greater scrutiny. We should, thus, be clear on exactly what is substituted in lieu of “revealed preferences”. The principal virtue of Posner’s “wealth maximization” or of revealed preference metrics is, for their other faults, their clarity. The substitution of revealed for corrected preferences necessarily requires a greater measure of normative justification or theoretical content to give it coherence and predictability.

What, then, is this content? The argument that misperceptions impel actors to positions contrary to their long-term interests often belies an ethics concerned with longevity. An individual with “distorted” pay-offs in smoking is, thus, said to better spread an increased measure of happiness over a longer period of time. His or her longevity represents the “true” metric to gauge the decision. Robert Frank, for example, expresses his commitments as

\[ \text{\textsuperscript{260}} \text{See Nozick, supra note 233.} \]

\[ \text{\textsuperscript{261}} \text{Sunstein and Thaler argue that market-preferences are often based simply on routine and custom rather than any objective measure and, accordingly, the government which tacitly allows extant routinized or customary ideational power to continue to form and determine market preferences is no less paternalistic than the one which seeks reform through intervention. Cass Sunstein and Richard Thaler, “Libertarian Paternalism Is Not an Oxymoron” University of Chicago Law Review 70:4 (Fall 2003) at 1194-1195. Whether we think of an interventionist government as paternalistic is a complex ideological question. It involves, for one, how deeply we commit to the dichotomy between public and private. Nevertheless, it is safe to say that, whether infused with extant power or not, the revealed preference metric is clearer.} \]
follows:262

By what criterion might we evaluate the efficiency of an individual’s choice of goals?... Darwin’s theory of natural selection, enriched to allow for the influence of cultural and other environmental forces during development, is an attractive candidate. In this framework, the design criterion for a goal or taste is the same as for an arm or a leg or an eye—namely, the extent to which it assists the individual in the struggle to acquire the resources required for survival and reproduction.

Human meaning is, thus, the maximization of the chances of survival, reproduction, and the preservation of genetic lines.263 Frank might not be speaking, here, for all behavioural economics.264 but his answer is paradigmatic. It is also rather one dimensional and hardly universal. The modern state is, no doubt, heavily oriented to the promotion of the health, vitality and virility of populations.265 This orientation is not unwavering. This concern for longevity or for the promotion of corrected preferences rests uneasily with liberal notions of autonomy, privacy and personal choice and freedom of contract. It is, moreover, in a clear tension with the principle that courts ought not to consider the providence of consensual bargains. The extent to which such commitments ebb and flow in the cases involves normative questions not reducible to a strictly “empirical matter”.266


263 Whether (or how) a concern for the longevity of the individual is connected to the posterity of a society is an interesting question. See Ronald Dworkin, supra note 226 at 200.

264 Berg and Gigerenzer, for example, suggest that it “remains an open question as to which departures from standard normative benchmarks in economics can be rationalized as adaptive”, for example, supra note 252 at 339.


266 Christine Jolls, Cass R. Sunstein, and Richard Thaler supra note 244 at 1545.
Conclusion

In this chapter I have tried to argue against the idea that insuring decisions could be used in something like an econometrics of law. Econometrics uses statistical methodology to test how economic theory bears out in the real world. Where the theory provides an inaccurate or incomplete picture, econometricians can suggest modifications to bring theory in line with the empirical evidence. If economic models of negligence are peopled with actors sharing uniform preferences and capacities, the actuarial data indicating otherwise presents an opportunity for theorists to question their assumptions. The diversity in an insurance pool is hardly trivial. Economists suggest that if differences were not taken into account by insurers in pricing, markets would collapse. It is easy to rebuild economic models or utility preferences to account for variety in the mutual pool. It is a harder question as to how this variety should be negotiated with law.

Variety presents a special problem for descriptive law and economics. Legal economists suggest that the shape of the law is generally structured, if not in form, in outcome, to produce economically efficient results. If economic actors are different, however, the law should optimally respond to these actors or classes of actors differently. This is not what the law usually looks like. It is seldom that the law affords a special credit to people who make more money and, thus, take a greater benefit from risk, for example, or who take a special, irregular satisfaction from courting a dangerous activity.

It might be true that in many instances the costs of segregating defendant classes into preference groups outweigh the economic benefits of more individualized and targeted laws. However, where these costs can be overcome, such laws would still seem to offend principles of equity and fairness that are hardly subordinate to efficiency commitments. Sometimes the law affords a credit to those lacking the capacity to make fully reasoned risk-decisions. And yet, efficiency theory would seem to mandate that an additional premium should bear on less-than-rational risk-takers to ensure that they subjectively internalize “true” consequences. This requires the weightier and less permissive standard. Again, this raises equity concerns.

The law simply accounts for diversity in ways that are much more complicated than can be accommodated in any one particular type of economic model. This puts the descriptive thesis into real doubt. While efficiency, wealth maximization or longevity, for example, certainly have an important place in the law of negligence or liability insurance, they share this place with principles of equity, proportionality, denunciation, restoration, distributive justice, autonomy and choice.

On the other hand, the problem of variety allows legal economists an opportunity to reflect on their prescriptive commitments. There is a great appeal in the idea that

\[ \text{\textsuperscript{268} Samuel Rea’s economic analysis of liability insurance law seems to concede the primary importance of compensation to how insurance arrangements are structured in law in, S. Rea “Economic Analysis of Insurance Law” International Review of Law and Economics, 13 (1993) 145 at 156: “Since the insureds will have limited wealth, the insurance provides compensation to the victim for the accident. Given that this compensation is not complete, the insurance may make the victims worse off if the losses caused by the greater frequency of accidents (which are undercompensated) dominate the additional compensation that is received from injurers with limited wealth. It is possible that the additional compensation dominates the value of the added risk.”} \]

governments ought to serve and work to maximize the public happiness. The evident variety in how people manage their risks presses economists to give workable and normatively justifiable meanings to the concepts of happiness they employ. The strict use of money as an absolute measure of happiness avoids the uncertainty of qualitative metrics. This, however, confers special privileges on those who possess it. It, moreover, relinquishes the claim to circumscribe the operation of markets in circumstances people might find repugnant. The use of a corrected preferences metric allows a broader literature to enter into the debate. On the other hand, it exerts a homogenizing influence on the seeming diversity of attitudes and approaches to risk and it empowers judges and experts\textsuperscript{270} to assert just what a corrected preference looks like. This chapter does not suggest which metric ought to be used or even to offer modest parameters for the debate. It would, however, urge that the negotiation of economic arguments into legal theory is inherently invested with normative suppositions which demand normative justifications. Such justifications will look nothing like the determinate, unfalsifiable proofs of economic theory.

\textsuperscript{270} Kahan, supra note 249 at 121-122.
Bibliography


Alfred P. Dorjahn, “Extenuating Circumstances in Athenian Courts”, *Classical Philology*,
25:2 (April 1930).


Leon Green, Rational of Proximate Cause (Kansas City: Vernon Law Book, 1927).


J.S. Mill, Utilitarianism, 4th ed. (London: Longmans, Green, Reader and Dyer, 1871)


