ECONOMIC RESEARCH REPORTS

THE IMPUTATION THEORY OF PROXIMATE CAUSE: AN ECONOMIC FRAMEWORK

Mario J. Rizzo

R.R. #81-18 February 1981

C. V. STARR CENTER
FOR APPLIED ECONOMICS

NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, N.Y. 10003

Contribution to Symposium
on Modern Tort Law, Georgia Law
Review

THE IMPUTATION THEORY

OF

PROXIMATE CAUSE

an economic framework

by

Mario J. Rizzo

February, 1981

THE IMPUTATION THEORY OF PROXIMATE CAUSE: An Economic Framework

by Mario Rizzo

ABSTRACT

A wrongfully starts a fire near B's barn. A very unusually strong wing arises and the fire spreads. B's barn is completely destroyed. Is A liable? This type of question is the subject matter of the theory of proximate or legal cause. This discussion paper shows the relevance of economics to such questions. By using a simple probabilistic marginal product approach, a theory of legal causation is developed. This theory attempts to explain most of the major doctrines in this area of law in a unified and coherent manner. A comparison is also made with Steven Shavell's alternative economic approach to causation.

In recent years, the economic efficiency hypothesis has attained recognition as an important development in the theoretical study of common law doctrines.[1] Unfortunately, many have come to identify this approach, with its attendant difficulties,[2] as the main or even sole contribution that economics can make to the analysis of common law. This, however, is far from accurate. The usefulness of economic tools is not restricted to the elevation of wealth maximization as the goal toward which the common law, particularly the law of torts, inexorably tends.[3] There is another approach that makes use of economic analysis which can contribute significantly to our understanding of tort law. This Article develops an economic imputation theory in the context of explaining major doctrines in the law of proximate cause. Our theory emphasizes the importance of tracing the relative objective marginal product of wrongful conduct in terms of the probability of that conduct bringing about the harm actually suffered.[4] In less formal lanaguage, the theory focuses on the degree of risk to which an instance of wrongful conduct exposes the plaintiff relative to that of other wrongful conduct or intervening events.

The specific thesis advanced and defended here is: a harm is proximately related to a wrongful act or omission if its objective marginal product in probabilistic terms is higher with respect to that harm than that of any other instance of wrongful conduct or intervening events. Consequently, according to the imputation approach, objective or best-informed probability, but not foreseeability, is important to the determination of proximate cause. What is likely or unlikely to the most perceptive man

("objective") need not be foreseeable or unforeseeable to the reasonable man. The economic efficiency analysis, on the other hand, implies that probabilities per se are irrelevant, but that foreseeability may, under certain circumstances, be crucial to establishing proximate cause.[5] Where optimal accident avoidance is the only goal of the tort system it is not at all puzzling that the degree to which something can be foreseen by the reasonable man is important. Individual expenditures on avoidance are linked to foreseeability and not to objective probability.

The sections that follow will reveal the great explanatory power of the imputation approach. The major proximate cause doctrines will be shown to be remarkably consistent with the implications of this theory. In sharp contrast, we shall see that the efficiency framework fails to explain some of the most elementary features of that body of law.

In section I the relationship between deterministic and probabilistic causal analysis is explored. Section II derives and develops the probabilistic theory of causation. Then, in section III, the theory is applied to direct causes and illustrated through the relevant case law. Section IV extends the theory to superseding and concurrent causation. Finally, in section V, the imputation and efficiency approaches are confronted in terms of their respective power to explain the rules[6] of proximate cause.

Part I: Theory

I. Deterministic and Probabilistic Causation

It is only from a formal point of view that any effort has to be expended in showing that deterministic causal analysis is merely a special case of the probabilistic. In everyday speech, causation entails both

determinist and probabilistic meanings.[7] When Professor X says "The food at the University Club causes me to get sick" he need not mean that he gets sick every time he eats there, or equivalently, that eating there necessarily makes him sick. It is sufficient that the club's food produces a "tendency" for gastro-intestinal discomfort. In fact, the occasions on which we clearly mean that X necessarily results in Y are no doubt far fewer than those in which we have some probabilistic conception in mind. The latter idea of causation clearly permeates our informal, everyday thought.

(A.) INUS Conditions. The formal framework within which our probabilistic theory of causation is developed is that of a modified[8] INUS condition.[9] This concept was originally advanced in a deterministic context but, as will be demonstrated later, is merely a special case of a more general probabilistic relationship. In simplest terms, an INUS condition is an insufficient but necessary part of an unnecessary but sufficient complex of conditions.[10] Suppose a number of longshoremen negligently dropped a plank into the hold of a ship. The friction produced by the impact of the plank against the hold caused a spark. Due to the presence of benzine vapor in the hold, the spark ignited a fire that completely destroyed the ship.[11] The wrongful conduct, A, can be an INUS condition if certain criteria are met. Let B, C, etc. be other conditions, both positive and negative, that are needed to form a minimally sufficient complex of conditions for the fire (Z). These conditions can be such things as the presence of benzine vapor and the absence of water in the hold (they can be summarized by "X").[12] Since A alone could not have produced the fire but the conjunction of conditions, AX, could, A is an insufficient moment in a sufficient (complex) condition. Although AX was not necessary for the fire (presumably it could have happened

in other ways), A was indeed necessary for the fire to happen in just the way it did. Thus, A is an insufficient but necessary moment in an unnecessary but sufficient condition. This, then, is very often what we mean when we say that A is a (the) cause of the ship's destruction.[13]

In more formal terms, we can say that AX is sufficient for the fire and that the disjunction[14] of AX and all other minimally sufficient conditions (Y) is both necessary and sufficient. Necessity and sufficiency are mere limiting cases of the probabilistic analysis. To say that AX is sufficient is equivalent to the statement that the probability of Z conditional on AX occurring is one. If (AX or Y) is necessary for Z then it follows that the probability of Z given the absence of the disjunction (AX or Y) is zero. The idea of an INUS condition can thus be fully stated as a special case in a probabilistic framework.[15] A is an INUS condition of the result Z if and only if:

(1) $P(AX \cup Y) > 0$ possibility[16]

(2) $P(Z|AX \cup Y) = 1$ sufficiency[17]

(3) $P(Z|\overline{AX} \cup Y) = 0$ necessity[18]

(4) $P(Z|A) = 0$ insufficiency

(5) $P(Z|X) = 0$ insufficiency

Condition (1) formalizes the requirement that there is more than one possible way for the event to occur (i.e., two or more minimally sufficient conditions). Conditions (2) and (3) specify the sense in which A is an element in a necessary and sufficient condition for Z. Finally, the last two conditions show that A and X are alone each insufficient for the overall result.[19] A cause can only produce a result in a given environment.

(B.) Causality as Functional Dependence. Another useful way to see the deterministic INUS condition as a subcategory of probabilistic analysis is to examine causality as functional dependence.[20] Consider the following probabilistic production function:

$$p(Q) = f(K,L).$$

This says simply that inputs K and L produce not a certain output Q but, rather, a given probability of a particular output level.[21] A farmer, for example, uses inputs to plant his crops but, because the weather is uncertain, he cannot be sure that his yield will be, say, a hundred bushels of corn. Instead, he might say that K and L produce a 0.5 probability of a hundred bushels of corn.[22] Suppose now that a and x are particular values of K and L. As such, they are sufficient conditions for a given value of p(Q). Although they are sufficient, they are, in general, not necessary since there will be some other combination of values that will produce the given p. Let y be the disjunction of all the other (minimal) sufficient conditions. All of the properties of the deterministic INUS condition framework can then follow. The event a (or x)[23] can be an insufficient but necessary moment in an unnecessary but sufficient complex condition for the determination of a given probability level.[24] We have simply transferred the determinateness from the outcome (Q) to the functional relation. The pure or conventional deterministic analysis is thus achieved as a special case when p equals one.

(C.) Particular Events vs. Classes. The two previous subsections have demonstrated how deterministic causal analysis can be seen as a subcategory in a probabilistic framework. Nevertheless, it might be thought that, even as a subcategory, deterministic causation is fundamentally

different from the more general probabilistic variety. The former, it is argued, is about particular events (e.g., this x is the cause of this z) while the latter is about classes of such events.[25] This, however, is a seriously incomplete view of the deterministic perspective. Both deterministic causal statements and probability statements[26] entail corresponding general propositions.[27] "X is necessary for Y" entails the counterfactual conditional "If not X, then not Y." Similarly, "X is sufficient for Y" entails the factual conditional "Since X, then Y." Both of these conditionals "are best considered as condensed or telescoped arguments" that have as premises "simple factual universals."[28] Thus, when we say that dropping this plank caused a fire to ignite in this ship's hold we are making an argument that involves several stages. First, there is a set of universal propositions about how in general sparks are created and benzine vapor ignites, etc. Second, there is a set of singular or particular statements about what is and is not the case in this instance. Finally, from the conjunction of universal and singular statements we can deduce the particular event. This is the canonical form of causal explanation.[29] Similarly, when we say that dropping planks in such and such a manner tends to cause fires to ignite we are making a similar argument. In this case, however, the premises are a set of propositions about the frequency with which sparks are created and benzine vapor is ignited under certain circumstances. Then, as before, there is a set of singular statements about this instance. Finally, the conjunction of the foregoing implies probabilistic conclusions about individual cases.[30]

A deterministic causal framework thus involves arguments with premises that are universal statements pertaining to classes of events. Similarly,

a probabilistic framework involves premises that are frequency generalizations about such classes. Hence the former are merely special cases of the latter generalizations where the relative frequency of the characteristic(s) equals one.[31]

II. A Probabilistic Theory of Causation

(A.) The Objective Marginal Product. If we bring together the idea of an INUS condition as the special case of a probabilistic relation[32] and the concept of causation as functional dependence,[33] it is possible to develop a thoroughly probabilistic INUS framework. A is a probabilistic INUS condition of the event Z if and only if:

(1) $P(AX^{u}Y) > 0$

(2) $P(Z|AX) = \beta < 1$[34]

(3) $P(Z|\overline{AX^{u}Y}) \neq \beta$

(4) $P(Z|A) < \beta$

(5) $P(Z|X) < \beta$

Each of the above requirements is analogous to those specified for the deterministic INUS condition. Requirement (1) expresses possibility as before. The second requirement should be interpreted that (AX) is minimally sufficient to produce a given probability level (β) for Z. It follows also that the disjunction of (AX) and all other minimal sufficient conditions (Y) is itself a sufficient condition.[35] Requirement (3) says that (AX or Y) is necessary for β. Hence it is both necessary and sufficient. Finally, requirements (4) and (5) make evident that neither A nor X are alone sufficient to attain the probability level β. In less formal

language, a probabilistic INUS condition, A, is an insufficient but necessary moment in an unnecessary but sufficient complex condition for the probabilistic result (i.e., a given probability level).

From the above framework we can derive the very useful idea of a "naive cause."[36] Requirements (2) and (5) can be written:

$$P(Z|AX) = \beta$$

$$P(Z|\bar{A}X) = \alpha < \beta$$

Therefore, A as a probabilistic INUS condition entails

$$P(Z|AX) > P(Z|\bar{A}X).$$

The event A at time to (A_{t_o}) is thus a naive cause of the event Z and t_1 (Z_{t1}) if and only if:

(1) t_o is earlier than t_1 ($t_o < t_1$)

(2) $P(Z_{t1} | A_{to}, X_{t_o}) > P(Z_{t1} | \bar{A}_{t_o}, X_{t_o})$.[37]

This says simply that one event is a naive cause of another if the presence of the former raises the probability of the latter. In this framework, a cause does not make its effect necessary but only more probable. A naive cause is a genuine cause only when it is not spurious.[38] An event A_{t1} is a spurious cause of Z_{t2} if and only if A_{t1} is a naive cause of Z_{t2} and there is a t_o ($\leq t_1 < t_2$) and A_{to} such that

(1) $P(Z_{t2} | A_{t1}, A_{to}, X_{t_o}) \leq P(Z_{t2} | \bar{A}_{t1}, A_{to}, X_{to})$

(2) $P(Z_{t2} | A_{t1}, A_{to}, X_{to}) > P(Z_{t2} | A_{t1}, \bar{A}_{to}, X_{to})$.[39] [40]

The idea of a spurious cause can be very simply illustrated. Suppose we notice that a crowded parking lot (A_{t1}) near the beach is positively correlated with a crowded beach (Z_{t2}).[41] Do we say that the crowded parking lot caused the beach crowd? Obviously not. There is a prior event, i.e., the heat (A_{to}), that causes both. So in this case, A_{t1} adds no

explanatory power to A_{t_o} and the probability of Z_{t_2} conditional on (or given) A_{t1} and A_{to} is no higher than it is conditional on A_{to} alone (requirement (1)). In fact, it may even be lower in the former case. This is because a crowded parking lot, given the heat, will make people less likely to go to the beach. Finally, if A_{to} is a genuine cause then it must add some explanatory power to A_{t1}, the spurious cause (requirement (2)).

In the rest of this Article we shall assume that a naive cause also meets the criteria for genuineness. Accordingly, we can define the following relationship:

$$\text{if } P(Z_{t1} \mid A_{to}, X_{to}) = \beta$$

$$P(Z_{t1} \mid \overline{A}_{to}, X_{to}) = \alpha < \beta$$

then $(\beta-\alpha)$ is the objective marginal product of A_{to} or, in other words, the increment in the objective[42] probability of Z_{t1} caused by A_{to}.

(B.) Raising the Probability of Harm. The previous subsection demonstrated that a probabilistic notion of causality entails the property that a cause must raise the probability of its effect. However, to say that A raises the probability of Z is an ambiguous statement. The context of such a claim must be specified carefully. Probability estimates are always relative to a background of particular facts and theories. In addition, it is important to make clear the state of affairs in relation to which the probability must be raised. As we shall see later, careless specification of the latter reference point is the source of many unnecessary confusions in the theory of proximate cause.

(i) The degree to which we depart from or approach the limiting case of deterministic analysis depends on the comprehensiveness of the information

set and theoretical framework that is used. Suppose we are aware of only some of the facts (conditions) prevailing at t_o, then even the very best of theories will enable us to predict the outcome at t_1 only probabilistically. On the other hand, if we were aware of all conditions at t_o, still no theory could perfectly link them together and predict the outcome deterministically. In reality, all deterministic models are ultimately illusory.[43]

In the framework developed here the relevant background of conditions and theories can be specified at two time periods. Our interest in the former is restricted to those prevailing[44] at t_o, the time of the conduct whose causal relevance we are tracing, but determined on the basis of the best information available at the trial. This is an "objective" assessment of the facts in the sense that we are not restricted to what the reasonable man could have known.[45] The theoretical framework used in conjunction with the facts to predict the outcome must also the best available at the time of the trial.[46] When we say that the reason X died of lung cancer in the nineteenth century was that he smoked too much, we are obviously not restricted to the medical theories of that century. Causal statements are objective in the sense that they are made with the best information at hand.

Now that the background of the probability statements used in this Article has been more precisely specified, the content of those statements ought to be examined. Although we have argued that causal analysis is perfectly consistent with a frequential view of probability,[47] it is not necessary to restrict ourselves to that perspective of probability. A great deal of modern research has accepted the position that probability is merely an informed degree of belief[48] and objective probability is the

best-informed degree of belief.[49] This modern view is far more useful for our purposes since it enables us to analyze cases that are not members of a large class with virtually identical fact patterns.

(ii) Determination of the states of affairs in relation to which probabilities must be raised is crucial to application of the imputation theory. In this framework we are interested in the causal relevance of wrongful conduct (e.g., A_{t_o}) in the context of standard environmental conditions.[50] Wrongful conduct is defined outside of our theory. Consequently this enables us to apply the analysis in both negligence and strict liability contexts.[51] The establishment of causal relationships need not be affected by the nature of the initial wrong itself. The standard environment (X_{t_o}) includes all those conditions that coexist wrongful conduct whether they are normal or abnormal as well as those subsequent events that are not "highly extraordinary."[52] Thus, the requirement of our theory is that wrongful conduct must raise the probability of the harm suffered relative to the probability that would have existed given the standard environment alone.[53] This is the relative marginal product of wrongful conduct.

The method of determining whether wrongful conduct actually raises the probability of a given harm can be seen most easily through the examination of a few examples.

(1.) Suppose an individual is wrongfully exposed to radiation and a year later dies of cancer. Is his death proximately caused by the wrongful conduct? The answer to this question depends on the risk of cancer created by the standard environment alone relative to the risk added by the exposure to radiation. If the probability that the particular individual would have

contracted cancer within the next year is .10 absent the wrongful exposure and remains at approximately the same level after the exposure (ceteris paribus), then the causation test has been failed.[54] The objective marginal product of wrongful conduct is zero.

(2.) Coincidence cases are the classic illustration of the absence of probabilistic causation. In Berry v. Borough of Sugar Notch[55] defendant had negligently permitted a tree to remain near the bus route. Plaintiff then drove a bus at a wrongfully fast speed along this route, passing under the tree just as it happened to fall. The negligence of plaintiff did not bar recover in this case because of the lack of causal connection between the harm and fast driving. In terms of the imputation framework, this can be expressed by saying that the probability of the harm (Z_{t1}) was no greater given the plaintiff's wrong (A_{to}) than it was absent that behavior and given only the defendant's wrong and the standard environment (here both can be compressed into X_{to}).[56] Thus,

$$P(Z_{t1} \mid A_{to}, X_{to}) = P(Z_{t1} \mid \overline{A}_{to}, X_{to}).$$

This, however, is an extreme case. In other situations the left hand side of the above relationship may exceed the right by a small amount. In these cases most of the causal relevance lies with the environment and not the wrongful conduct and hence, without apportionment, plaintiff should not recover at all.[57]

(3.) The final example that we shall consider is that if "alternative causation."[58] Suppose B is about to eat food containing some ingredients to which he is highly allergic. If we were to eat this at t_o then at t_1 B would have a 0.5 probability (or higher) of a fatal stroke. However, A wrongfully frightens B at t_o (A_{to}) which prevents him from eating. As a

consequence of the excitement and fear, B's probability of a fatal stroke at t_1 is 0.5. Is A's conduct proximately related to B's actual fatal stroke (Z_{t1}) if that should come to pass? The answer is yes. At t_1 only one minimally sufficient condition for the 0.5 probability actually existed and A's frightening B was a necessary moment in that condition. The allergy-producing food, however, was not. Hence the former was "necessary post-factum" for the probabilistic result.[59] In other words, A's conduct was necessary, as things actually developed, to raise the probability of a fatal stroke beyond what it would have been given the standard environment alone. Thus, $P(Z_{t1} | A_{to}, X_{to}) > P(Z_{t1} | \bar{A}_{to}, X_{to})$ because X_{to} is the same in both cases.

(C.) The Defined Consequence. When we analyzed wrongful conduct as raising the probability of a given harm, we left the term "harm" largely undefined. Unfortunately, there is considerable ambiguity attached to that term. The purpose of this subsection is to specify more clearly what is meant by "harm." The consequence with which, by definition, we are concerned is not merely the resulting pecuniary loss but rather the "loss-creating event"[60] itself. To see the precise difference between the two consider the following example. Suppose that a carefully-performed operation carries a .01 risk of blinding in one eye and this involves a pecuniary loss of $50,000. If the same operation were negligently-performed (e.g., by working too quickly) there would be no risk of blinding since that requires the patient to be under an anasthetic for a longer time. However, there would now be a .01 chance of paralysis in one leg ($50,000 associated loss) and a .02 risk of death as well. If the doctor operates negligently and the patient becomes paralyzed, can we argue that this is a coincidence? In

both the nonnegligent and negligent cases there is a .01 probability of the same pecuniary loss. Wrongful conduct did not raise the probability of the pecuniary loss that actually came about (and hence has a zero marginal product in pecuniary terms). The law answers our question in the negative. Only when "lawful conduct would have brought about the harm in a manner substantially indistinguishable in point of time, place, manner and detail of occurrence" from what, in fact, happened can we say that no causal connection exists.[61] Causal arguments are constrained to what actually did occur and not what might have. Hence it is no defense to say that if the negligent act hadn't produced a paralyzed leg a substitute careful act might have produced a blinded eye.[62]

Part II: Applications

I. Direct Causation: A Framework

(A.) Simple Direct Causes. Direct causation is very prominently treated in the literature on proximate cause.[63] In this subsection we shall formalize this idea within the context of the imputation theory. The formalization clarifies the law's underlying logic in holding that the absolute level of risk creation is irrelevant for causal connection in these cases. Wrongful conduct, A_{t_o}, is a simple direct cause of the harm, Z_{t_2}, if and only if: (i) A_{t_o} is a naive cause of Z_{t_2}; (ii) there is no intervening event (I_{t_1});[64] or (iii) if there is an intervening event, then $P(Z_{t_2} \mid A_{t_o}, X_{t_o}) \cong P(Z_{t_2} \mid A_{t_o}, I_{t_1}, X_{t_o})$.[65] The third requirement says simply that although an event has intervened between the original wrong and the ultimate harm, it is completely noncausal. I_{t_1} adds nothing to the

explanatory power of A_{t_o} or, in other words, to the overall probability of the harm given A_{t_o}.

There is nothing in the definition of a simple direct cause requiring that the probability of harm conditional on wrongful conduct be high (i.e., that the harm be a "foreseeable" consequence). The only requirement is that it "substantially" exceed the probability conditional on $\overline{A}_{t_o}$ and X_{t_o}. Let $P(Z_{t_2} | \overline{A}_{t_o}, X_{t_o})$ equal α and $P(Z_{t_2} | A_{t_o}, X_{t_o}) = \beta$, then causal proximity is established if the objective marginal product of A_{t_o} $(\beta-\alpha)$ is greater than the marginal product of the standard environment (α). Under these circumstances, since the preponderance of causal responsibility lies with the defendant, it is reasonable to place liability on the defendant.

The consistency of direct causation with a low conditional probability of harm is related to the limitations of our knowledge. Although we may not be able to predict very well (due to ignorance of relevant conditions or theories), the relationship between A_{t_o} and Z_{t_2} provides, in the present state of knowledge, the best explanation available.[66] In other words, the instant wrongful conduct explains better than the "unaided" standard environment. Since the court must operate within the constraints of its knowledge, it is only the relative marginal products that matter and not their absolute levels.

(B.) Direct Causes with Dependent Intermediate Causes. Analytically, this case is not substantially different from that discussed above. Indeed, in every instance of direct harms there may be some intermediate causes if the event is described in fine enough detail. However, because some cases have a very pronounced "domino" effect, it appears useful specifically to treat this variation.

A_{t_o} is a direct cause with a dependent intermediate cause(s) (M_{t_1}) if and only if: (i) A_{t_o} is a naive cause of Z_{t_2}; (ii) $P(M_{t_1} | \bar{A}_{t_o})$[67] = α is low and hence M_{t_1} is not part of the environment; (iii) $P(M_{t_1} | A_{t_o}) = \beta$ and $(\beta-\alpha) > \alpha$; and (iv) $P(Z_{t_2} | A_{t_o}, M_{t_1}) > P(Z_{t_2} | A_{t_o}, \bar{M}_{t_1})$. Two parts of this definition require some explanation. Conditions (ii) and (iii) establish the dependency of the intermediate event. If we view α as the marginal product of the standard environment in producing M_{t_1}, then $(\beta-\alpha)$, the increment in probability added by A_{t_o}, is the marginal product of A_{t_o} with respect to M_{t_1}. When the marginal product of the latter is greater than the former it seems reasonable to call M_{t_1} "dependent." Condition (iv) establishes the causal relevance of M_{t_1}: it adds to the explanatory power of A_{t_o}. Nevertheless, at t_o the single most satisfactory explanation for Z_{t_2} is A_{t_o} rather than the standard environment. Suppose, for simplicity, that M_{t_1} gives rise to Z_{t_2} with probability equal to one, then, since $P(M_{t_1} | A_{t_o}) > P(M_{t_1} | \bar{A}_{t_o})$, A_{t_o} must be more explanatory, i.e., its marginal product exceeds that of any rival. In this case, the marginal product of wrongful conduct exceeds that of the standard environment and hence the proximity of causal connection between A_{t_o} and Z_{t_2} is established.

This variety of direct causation can illustrate the difference between improbable consequences of wrongful conduct and the intervention of improbable events.[68] If $P(M_{t_1} | A_{t_o})$ is not one but is still high (e.g., 0.7) and there are many such domino stages (M_{t_2} ... M_{t_7}) with the same conditional probablities, then the probability of the overall harmful result will be rather small (e.g., slightly more than 0.08). Here, because of the compounding of probabilities at each stage, there is an improbable consequence without improbable intervention. As we have already seen,

improbable consequences per se are irrelevant for the determination of proximate cause. Improbable interventions, on the other hand, can negative causal connections. This issue will be explored subsequently.[69]

II. Direct Causation: Illustrations

In this subsection our hypothesis about the proximity of direct causal relationships will be tested by reference to a representative sampling of the case law. Recall the imputation approach implies that when the objective marginal product of the instant wrongful conduct exceeds that of the standard environment, the former is causally linked to the given harm. The objective (or even reasonably apprehended) absolute likelihood of harm is irrelevant; only relative marginal products are important.

(1.) In re Polemis.[70] Earlier in this Article the nature of a deterministic INUS condition was illustrated by reference to the facts of this famous case.[71] However, Polemis can also serve as a straightforward example of simple direct causation in our probabilistic framework. The dropping of a plank in the hold of a ship was negligent because, as the arbitrators found, some harm (e.g., a dent in the hold) was foreseeable. The actual harm, however, the destruction by fire, was not similarly foreseeable. The court, nevertheless, held the defendants fully liable because the outcome was the "direct" result of their negligence. Foreseeability is irrelevant in cases of direct causation.[72]

Assume that the defendants were responsible for the escaped benzine vapor that caused the spark to ignite a fire. Hence their negligence (A_{t_o}) consisted in both dropping the plank and creating a dangerous condition. We can than compare the probabilities of the ship's destruction (Z_{t_1}) in the

following way. If $P(Z_{t_1} | \bar{A}_{t_o}) = \alpha$ and $P(Z_{t_1} | A_{t_o}) = \beta$ then, as seems likely, $(\beta-\alpha) > \alpha$. In other words, the marginal product of the wrongful conduct exceeds that of the "unaided" standard environment.[73] Thus, A_{t_o} is a proximate cause of Z_{t_1} despite the apparently low absolute level of β ("unforeseeability").[74]

(2.) Flooding Hypothetical. In the previous case there was no need to consider the causal status of an intervening event because of the assumption that the defendants were responsible for the benzine vapor. The following is an example of intervention that lacks any causal significance. Suppose defendant defectively constructed a drainage system. Soon after this, an abnormally heavy rainstorm arose and an area of town became flooded. As a consequence of the flooding a valuable art object in a nearby home was completely destroyed. Is the harm proximately related to the wrongful drainage construction? If a normal rainfall were just as likely to destroy the object, then the case is still a direct harm. Although there was an intervening event (the abnormal storm) it is totally noncausal. Since the marginal product of that event is zero (i.e., it added no risk to that already created by the wrongful construction), defendant's negligence is the proximate cause of the artwork's destruction.[75]

(3.) Hill v. Windsor.[76] The final example in this section is a classic "domino" case in which the principles of direct causation are not in the slightest degree altered by the existence of dependent intervening events. Defendant's tug boat negligently bumped into the fender of a bridge. The force of that collision was transmitted through a number of piles and eventually caused a plank held between two piles to fall out. This, in turn, resulted in plaintiff getting his leg caught between the piles as they sprang together. Defendant was held liable.

Let A_{t_o} = the original negligence; M_{t_1} = the conjunction or vector of all intermediate stages, and Z_{t_2} = the accident. To test the dependency of the intermediate events we compare, as before, $P(M_{t_1} | \bar{A}_{t_o}) = \alpha$ and $P(M_{t_1} | A_{t_o}) = \beta$. If $(\beta-\alpha) > \alpha$ then M_{t_1} is dependent. Assume, as is extremely likely, that M_{t_1} adds explanatory power to A_{t_o}. Thus, $P(Z_{t_2} | A_{t_o}, M_{t_1}) > P(Z_{t_2} | A_{t_o}, \bar{M}_{t_1})$. M_{t_1} is therefore a dependent intermediate cause with respect to Z_{t_2}. From this it follows that the original wrong is the proximate cause of plaintiff's injury. The marginal product of A_{t_o} in producing M_{t_1} is greater than that of the standard environment, and the marginal product of M_{t_1} with respect to Z_{t_2} is, by assumption, no different given A_{t_o} or $\bar{A}_{t_o}$. Hence the marginal product of A_{t_o} with respect to Z_{t_2}, although doubtless quite small, is greater than that of the environment.[77] When the marginal product of wrongful conduct exceeds that of its rivals, the former is the proximate cause of the harm.[78]

The above cases illustrate a general rule: when a harm is the direct outcome of wrongful conduct, i.e., when no independent cause intervenes, liability follows even though the result may have been unforeseeable.[79] From time to time, exceptions to this rule have been suggested.[80] The most notable of these is derived from the claim that "[t]here must be obviously some point along the chain of direct causation, in the sense of one force activating another, where the idea of liability is repugnant to one's conception of basic justice."[81] Although we are not told what this conception of basic justice is, it appears that in a small handful of cases, when the result is extraordinarily freakish, liability has been denied.[82] However, many of the cases in which "foreseeability" appeared to be at issue in the denial of liability were not, in fact, cases of direct causation.

Some event or voluntary behavior intervened between the original wrong and the harm.[83] Furthermore, although one of the most famous foreseeability cases, Palsgraf v. Long Island R.R.,[84] was almost a direct harm, the defendant sued was not responsible for the main causative agency. In effect, the source of the far less important intervening conduct was sued. Thus, the result can be explained on causal grounds alone without recourse to the foreseeable plaintiff issue.[85]

III. Superseding Causation: A Framework

(A.) Superseding Events. When intervening events have considerable independent causal significance they often break the link between an initial wrong and an ultimate harm. The event S_{t_1} is a superseding cause with respect to wrongful conduct, A_{t_o}, if and only if: (i) A_{t_o} and S_{t_1} are naive causes of Z_{t_2}; (ii) $P(S_{t_1}) \tilde{=} 0$ or, at least, is very low (abnormality); (iii) $P(S_{t_1} | A_{t_o}) = P(S_{t_1} | \overline{A}_{t_o})$ (independence); and (iv) $P(Z_{t_2} | A_{t_o}) = \alpha$ and $P(Z_{t_2} | A_{t_o}, S_{t_1}) = \beta$ such that $(\beta-\alpha) > \alpha$. The requirements of this definition are not arbitrary but follow immediately from our basic theoretical perspective. A superseding cause must be a low probability event to distinguish it from the standard environment. If the probability of S_{t_1} at t_o is very small then the marginal product of A_{t_o} cannot incorporate the effect of S_{t_1} except to a negligible extent.[86] In this case, therefore, the risk-creation of S_{t_1} must enter the picture separately. In addition, a superseding cause must be independent of the original wrongful conduct so as to avoid the argument that it is merely the result of the conduct it allegedly supersedes. Otherwise, it is merely a case of direct causation with dependent intermediate causes. As we saw previously, these causes do not interrupt the chain of

causation: their effects are imputed backward to the original wrongful conduct (causa causans).[87] Finally, a superseding cause must have a probabilistic marginal product exceeding that of A_{t_o}. Suppose A_{t_o}, in conjunction with the standard environment, raises the probability of Z_{t_2} by α and when S_{t_1} is conceptually added to the scenario the probability of the harm rises to β. S_{t_1} supersedes A_{t_o} if the risk added by the former $(\beta-\alpha)$ exceeds that of the latter (α). When the system is constrained not to apportion,[88] it makes sense to place liability on wrongful conduct only if its causal significance (marginal product) is greater than that of the intervening event. Otherwise, the intervening event supersedes and plaintiff must shoulder the burden. Here again relative marginal products determine the proximity of causal connections.

(B.) Superseding Wrongful Conduct. This case bears some similarity to the previous framework for superseding events. Nevertheless, there is an important difference between the two that requires elaboration. B_{t_1} is a superseding wrongful cause with respect to A_{t_o} if and only if (i) both A_{t_o} and B_{t_1} are naive causes of Z_{t_2}; and (ii) $P(Z_{t_2} \mid A_{t_o}) = \alpha$ and $P(Z_{t_2} \mid A_{t_o}, B_{t_1}) = \beta$ such that $(\beta-\alpha) > \alpha$. Here the unconditional and conditional probability at t_o of subsequent wrongful conduct is irrelevant.[89] Recall that our central hypothesis is that proximity of causal relationships is determined by the relative marginal products of wrongful conduct. Hence subsequence wrongful conduct can never be part of the standard environment regardless of how likely it is. The prior determination of the given activity as wrongful means that it is potentially eligible to bear liability. Whether it bears such liability is determined, ex hypothesi, by its relative marginal product and this cannot be measured if it is buried in the standard environment.[90]

In our framework concurrent causation is a special case of the superseding causal analysis. If $(\beta-\alpha) \cong \alpha$ then the two wrongful causes are concurrent.[91] Neither can be said to dominate the other and hence the harm is truly a joint product of the two. Since there is no reason to place liability on one or the other defendant, the common law generally makes them joint and severally liable for the whole harm.[92]

(C.) Simultaneous Wrongful Causes. A_{t_o} and C_{t_o} are concurrent simultaneous wrongful causes of Z_{t_1} if and only if (i) both A_{t_o} and C_{t_o} are naive causes of Z_{t_1} and (ii) $P(Z_{t_1} | A_{t_o}) = P(Z_{t_1} | C_{t_o})$. Simultaneous causes differ in an important way from either concurrent or superseding causes that are separated in time. The very fact of simultaneity makes it impossible to ask the question whether A_{t_o} added considerable risk to C_{t_o} or vice versa. Neither precedes the other temporally or logically. Hence we can only compare the risk created by each alone relative to the standard environment (since the environment clearly precedes them). When each cause, ceteris paribus, has roughly the same marginal product they are concurrent. Neither is more causally responsible for the harm than the other and so joint and several liablity prevailed at common law. On the other hand, if the marginal product of one clearly exceeds that of the other, then the former "supersedes"[93] the latter.

IV. Superseding Causation: Illustrations

(A.) Interventions: Unforeseeable or Abnormal? As we have seen,[94] if a subsequent event is highly improbable then it will not be considered part of the standard environment. At the time of wrongful conduct a very

low probability subsequent event will virtually not affect the marginal product of that conduct. Since our theory is based on the objective marginal product of conduct and events, the perspective from which the improbability of the intervention ought to be judged cannot be simply that of the reasonable man. The subsequent event must be abnormal and not (merely) unforeseeable: its objective probability must be low.

Although the language of the cases is often in terms of the foreseeability of interventions, the word "foreseeability" takes on a different meaning in this context.[95] According to the Restatement, the intervening event must appear "to the court in retrospect that it is highly extraordinary."[96] More specifically, it states:

> "In advance, the actor may not have any reason to expect that any outside force would subsequently operate and change the whole course of events from that which it would have taken but for its intervention. None the less, the court, knowing such a force has intervened, may see nothing extraordinary either in its intervention or in the effect which it has upon further development of the injurious results of the defendant's conduct."[97]

The idea of retrospective calculation of probabilities can be puzzling at first since, ex post, either something has occurred or it has not. A moment's reflection, however, reveals that the Restatement is referring to an objective assessment of likelihood based on all the facts existing at t_o (whether the defendant knew of them or not)[98] in conjunction with the best theoretical-causal framework available. The law's treatment of intervening events is therefore consistent with the objective imputation

approach developed in this Article. It is not consistent, as we shall see,[99] with the efficiency theory of causation.

(B.) Case Law. In this subsection we analyze three important cases from the perspective of the imputation theory. The analysis produces results consistent with the actual outcomes of these cases but enables us to reach those results in a much simpler and more consistent fashion.

(1.) Toledo & Ohio Central R.C. v. Kibler & Co.[100] The facts in this case provide a classic example of superseding causation in a probabilistic framework. A negligent delay by the railroad caused some goods that were in transport to be in a location where an extraordinarily unusual flood occurred. The goods were then destroyed by the flood. The court found that the negligence of the defendant, although a necessary condition, was not a proximate cause of the harm. The analysis on the imputation theory is straightforward. The probability of such destruction of the goods consequent upon negligent delay is surely very small. However, once the train is in the area of the flood, the objective probability of the accident rises dramatically. Thus the flood is a superseding cause with respect to the delay.[101]

(2.) Palsgraf v. Long Island R.R. Co.[102] Recall that one wrongful cause supersedes another if the risk added by the former exceeds that added by the latter. Normally, the superseding cause follows the other but the same principle clearly applies regardless of the temporal sequence. An earlier cause will supersede in importance a later one when the relative marginal product of the second is low. Palsgraf is a classic illustration of this situation.

A man carrying a package was running to catch a train as it was moving out of the station. He caught up to it "jumped aboard the car, but seemed unsteady as if about to fall"[103] when two guards negligently helped him. One pushed him from the rear while the other grabbed him up. In the process, the package he was carrying became dislodged. Unbeknownst to the trainmen, the package contained fireworks and exploded upon impact to the rails. Mrs. Palsgraf, who was standing some distance away, was injured when some scales fell on her as a result of the explosive shock. Cardozo's decision was that she could not recover against the railroad because they were not negligent as to her: she was an unforeseeable plaintiff. The case, however, can be easily analyzed on the imputation theory with the same result in a way that shows: (a) that the person responsible for the main causative force was not sued, and (b) the general irrelevance of foreseeability.

The probability (α) that Mrs. Palsgraf would have been harmed by falling scales given only the standard environment was plausibly negligible. In this context, the man carrying the explosives created a dangerous condition that alone could have caused the harm with a higher probability (β). Intervening between the dangerous condition and the falling of the scales was the "assistance" of the trainmen. This, together with the dangerous condition, raised the overall probability to γ. On the record of the facts, it seems likely that β is very close to γ since the man "seemed unsteady as if about to fall." However, when Mrs. Palsgraf sued the railroad, the relevant issue is the causal contribution of the trainmen in the context in which they acted. Since $\beta \cong \gamma$ that contribution is very small. On the other hand, it seems likely that $(\beta-\alpha) > \alpha$ and so if she had sued the creator of the dangerous condition, she would have recovered. The man with the explosives

is the proximate cause of the harm and not the railroad, the relative marginal product of which is low. Hence Palsgraf is a case in which the defendant actually sued was not the main causal agent: the original dangerous condition superseded in causative importance the trainmen's intervention. The issue of foreseeability is thus merely a roundabout way of arriving at a result more easily derived from this simple fact.[104] On the other hand, if the man most likely would not have fallen on his own, then $\beta \cong \alpha$ and $(\gamma-\beta) > \beta$. Thus, the railroad would have been a superseding cause.

(3.) Anderson v. Minneapolis R. Co.[105] In cases of simultaneous interaction of causes, whether both are wrongful or one is wrongful and the other innocent, courts frequently apply the "material" or "substantial" factor rule.[106] In the instant case two independent fires combined to destroy plaintiff's property. It was held that unless defendant's fire was a "substantial" factor in producing the harm it was not a proximate cause. Although the court did not interpret the words "substantial" or "material," it is probable that they refer to the general harm-creating propensity of each fire. If in a particular case each fire was sufficient to produce the harm, then it is only by at least implicit reference to a general class of cases that the substantiality of a cause may be determined. One fire will be a substantial cause relative to another if the former, taken alone, is typically or, on average, more likely to produce the harm than the latter. Thus, if a very large fire joined a much smaller one to destroy plaintiff's property, the former alone would be proximately connected to the harm on this rule.[107] In other words, the more the marginal product of the former exceeds that of the latter, the more likely the court will

consider the first a substantial factor.[108]

V. Imputation versus Efficiency

The imputation theory developed here takes the meaning of wrongful conduct as determined exogenously and then proceeds to trace the objective marginal product of that conduct. In a negligence framework, for example, failure to take care is the event whose marginal product is of concern. Under strict liability, on the other hand, the imputation method is applied to the activity itself which gives rise to the harm regardless of the level of care undertaken. We have seen that the improbability of a harm or the abnormality of the conditions coexisting wrongful conduct are irrelevant to the determination of proximate cause either in our theory or in current law. The objective marginal product of wrongful conduct is obviously affected by all coexisting conditions whether the defendant is mindful of them or not. On the other hand, the improbability of a subsequent event can negate the causal connection between an act or omission and the harm. This also follows from our framework. The objective marginal product of a wrong at t_o can only be neglibly affected by remote possibilities.

In this section it will be demonstrated that there are at least two grave flaws in the efficiency explanation of the fundamental proximate cause doctrines. First, the defined events are improperly specified. In a strict liability framework, the efficiency theory requires concern with the marginal product of taking due care while, ironically, in a negligence framework it is totally unconcerned with the marginal product of care. Secondly, the efficiency theory cannot explain both the law's treatment of coexisting abnormalities -- improbable consequences and subsequent

abnormalities. Either it explains the former but not the latter or vice versa. The imputation theory, as we have seen, can explain both.

(A.) Defined Event Problem.

(i.) Strict Liability. In this case, the efficiency theory requires that ". . . the scope of liability . . . be restricted . . . to types of accidents . . . that, given the circumstances under which the type of accident occurs, the effect of an increase in care in reducing accident losses should be sufficiently pronounced"[109] relative to the administrative costs of imposing liability. This creates incentives to take the proper level of care while economizing on administrative costs.[110] It does not pay to impose liability when the marginal product of taking care is less than the cost of achieving a specified level of care. Although Shavell's economic analysis is correct, it is clear that this paints a picture of proximate cause that is thoroughly at odds with existing law. The determination of causal connection under strict liability has nothing to do with the degree to which failure to take reasonable care would yield substantial benefits in terms of accident prevention. It does involve, however, the degree to which engaging in the activity itself raises the probability of harm.[111]

(ii) Negligence. In order to create the proper incentives to take care while minimizing the administrative costs of liability imposition, Shavell shows that "the sole criterion for including an accident in the scope of liability should be the size of accident losses."[112] Under a negligence system, defendant does not pay for the cost of nonnegligent accidents. Therefore, the private saving from exercising due care is the entire amount of losses that occur if an individual is negligent. Since, in general, only some of these are avoidable with the exercise of due care,

the private saving is greater than the social. Hence we can induce people to take proper care by restricting liability to the states of the world (i.e., given circumstances) in which absolute accident losses are large, even if the marginal product of taking care is low. By excluding cases in which accident costs are low administrative costs are reduced while only negligibly affecting the incentive to take care.

This, again, is inconsistent with generally accepted proximate cause doctrines. Under negligence, causal connection must be demonstrated between the wrongfulness of defendant's conduct (i.e., his failure to take care) and the harm suffered by palintiff.[113] This is typically interpreted to mean that if care would not have materially affected the harm in the given circumstances, then defendant's wrong is causally irrelevant.[114] In other words, the marginal product of care must be significant. In Shavell's system, on the other hand, it is possible that a state of the world would be included in the scope of liability in which failure to adhere to the due care standard was not a cause-in-fact of the accident.[115]

(B.) The Role of Foreseeability.

(i.) Correct Probability Estimates. The efficiency criterion for inclusion of an accident in the scope of strict liability (i.e., a relatively high marginal product of care) is valid irrespective of the low probability attached to the circumstances under which that accident occurs.[116] Inclusion of accident costs under improbable states of the world has, it is true, only a small effect on the incentive to take care since the expected[117] marginal product of care is raised only slightly.[118] On the other hand, the increase in costs associated with liability imposition will also be discounted by the probability of that state of the world.

Since the same probability discount appears on both sides of the benefit-cost ledger, we can disregard it and compare directly the marginal product of care with administrative costs.

This can explain the thin skull rule and, in general, the law's treatment of coexisting abnormalities. However, it runs contrary to the legal view of improbable subsequent events.[119] Under the efficiency view the occurrence of a given intervening event is analytically identical to the occurrence of any other improbable state of the world. In other words, one implication of the efficiency model is that coexisting and subsequent abnormalities ought to be treated identically: neither ought to negative causal connection.

(ii.) Underestimated Probabilities. If people systematically underestimate low probability events,[120] then the improbability of a subsequent event can be important.[121] Now if a state of the world is excluded from the scope of liability, the expected value of taking care would be reduced by less than in the previous case due to the underestimation. Expected administrative costs, on the other hand, would correctly be reduced by the same amount as before. Hence the expected costs of liability exclusion in terms of diminished incentive to take care are less than previously while the expected benefits are the same.[122] Thus, depending on the degree of underestimation, it would be consistent with efficiency to conclude that low probability events should negative causal connection. This argument, however, proves too much. In contrast to the previous case, the underestimation model can explain the law's view of subsequent abnormalities but not of those abnormalities coexisting wrongful conduct. From the efficiency perspective, they are both analytically identical since under-

estimation of either will have the same diminishing effect on the incentive to take care. Subsequent abnormalities are now explained at the cost of "unexplaining" coexisting abnormalities. The efficiency theory cannot have it both ways.

Concluding Remarks

This Article has shown that traditional deterministic ideas of causation are simply special cases of a more general probabilistic analysis. A probabilistic theory of causation coupled with the economics of marginal product imputation can explain major doctrines in the law of proximate cause. Thus, the usefulness of economics to understanding the law need not be restricted to a world inhabited by the Cult of Wealth Maximization. It is possible to use economic tools in a way that does not do violence to the basic goals and methodology of the common law.

FOOTNOTES

* Assistant Professor of Economics, New York University (on leave); Fellow in Law and Economics, University of Chicago Law School. This is part of a general investigation into the relationship between economic and legal reasoning for which The Spencer Foundation, through New York University, has provided financial assistance. I am also indebted to the Fred C. Koch Foundation, Pfizer, Inc., and the Scaife Foundation for support of my research.

1. For an excellent example of this approach see R. Posner, Economic Analysis of Law 27-161 (2d ed. 1977) and the sources cited therein.

2. See, e.g., Rizzo, Uncertainty, Subjectivity and the Economic Analysis of Law, in Time, Uncertainty and Disequilibrium 71 (M. Rizzo ed. 1979); Rizzo, Law amid Flux: The Economics and Negligence and Strict Liability in Tort, 9 J. Legal Stud. 291 (1980); Rizzo, The Mirage of Efficiency, 8 Hofstra L. Rev. 641 (1980). The moral and philosophical aspects are analyzed in Dworkin, Is Wealth a Value?, 9 J. Legal Stud. 191 (1980); Kronman, Wealth Maximization as a Normative Principle, 9 J. Lelag Stud. 227 (1980).

3. Even Landes and Posner have their doubts that there is any clearly defined process in common law adjudication that pushes the system

toward efficiency. See Landes and Posner, Adjudication as a Private Good, 8 J. Legal Stud. 235, 259-284

4. For an analysis of this principle in the context of damage apportionment see Rizzo and Arnold, Causal Apportionment in the Law of Torts: An Economic Theory, 80 Colum. L. Rev. 1399 (1980).

5. See Shavell, An Analysis of Causation and the Scope of Liability in the Law of Torts, 9 J. Legal Stud. 463, 482-484, 490-492 (1980). See also R. Posner, supra note 1 at 130 (importance of foreseeability limitation).

6. Thus, this Article is in fundamental disagreement with Judge Andrews famous remark in his dissent to Cardozo's decision in Palsgraf. Proximate cause "is not logic. It is practical politics." Palsgraf v. Long Island R.R., 248 N.Y. 339, 352, 162 N.E. 99, 103 (1928).

7. See P. Suppes, A Probabilistic Theory of Causality 7 (1970) and Rosen, In Defense Of A Probabilistic Theory Of Causality, 45 Philosophy of Sci 604, 612 (1978) (agreement with Suppes' view).

8. The Hart and Honoré distinction between causes and mere conditions (environment) is added. See H.L.A. Hart and A. Honoré, Causation in the Law 30-41, 67-76 (1959) [hereinafter cited as Hart and Honoré].

9. See Mackie, Causes and Conditions, 2 Am. Phil. Q. 245 (1965); see also Marc-Wogau, On Historical Explanation, 28 Theoria 213 (1962) (a similar theory).

10. See Mackie, supra note 9 at 245-247.

11. In re Polemis & Furness, Withy & Co., [1921] 3 K.B. 560.

12. Each element or moment in the sufficient condition can be equally called the cause. See Mackie, supra note 9 at 253. However, we add the causes and mere conditions distinction. See Hart and Honoré, supra note 8 at 67-76.

13. Neither the deterministic nor probabilistic version of the theory conveys everything we mean by causation. The requirements of an INUS condition are a very important aspect of causation. "I suggest that when we speak of the cause of some particular event, it is often a condition of this sort that we have in mind." Mackie, supra note 9 at 245.

14. By "disjunction" is meant the set of events (AX or Y), i.e., AX and/or Y. See, generally, R. Jeffrey, The Logic of Decision 55-56 (1965).

15. See Suppes, supra note 7 at 76. The fourth and fifth conditions, however, have been altered. See infra note 19.

16. The notation here signifies the probability of the union (U) of AX and Y, i.e., the set of all elements (events) belonging to AX or to Y or to both. This is the symbol for the word "disjunction." See supra note 14.

17 This means the probability of Z given (or conditional on) the union of AX and Y.

18. The line over $AX^{u}Y$ means not $(AX^{u}Y)$ or the absence of both AX and Y.

19. Suppes states these conditions as merely < 1; however, this is not an accurate rendering of the idea of an INUS condition. In a deterministic framework, insufficiency must mean that it cannot happen given only A or X.

20. See, generally, Mackie supra note 9 at 26-261 and Wold, Causality and Econometrics, 22 Econometrica 162, 165-166 (1954).

21. A formulation more consistent with the Hart and Honoré dichotomy between the cause(s) and mere conditions might be to consider, say, L an "environmental variable" and write

$$p(Q) = f(K;L)$$

where L is exogenous to the problem at hand. For the economics of environmental variables, see R. Michael, The Effect of Education on Efficiency in Consumption 7-11 (1972).

22. The alternative of saying that there is a probability

distribution over many different levels of Q is not useful here. In causal analysis we are constrained to what actually did happen. Since causes of action arise ex post we know the outcome.

23. See supra note 12.

24. (ax or y) is both necessary and sufficient for p(Q) while neither a nor x are sufficient alone.

25. See Borgo, Causal Paradigms in Tort Law, 8 J. Legal Stud. 419, 425 (1979).

26. Here attention restricted to the relative frequency interpretation of probability. But see text accompanying note 48, infra.

27. See Mackie supra note 9 at 253-254.

28. Id. at 254.

29. See K. Popper, The Logic of Scientific Discovery 60 (1958).

30. This assumes that when we knew nothing except that the instant case is a member of a class with a certain relative frequency of outcomes we can apply that relative frequency to the individual case. See, generally, H. Reichenbach, Experience and Prediction 312-319 (1938).

31. We need not, however, interpret probability statements frequentially. If we do not and adopt, say, the informed degree of belief interpretation then probability statements are also about particular events. See, e.g., DeFinetti, Foresight, Its Logical Laws, Its Subjective Sources in Studies in Subjective Probability (H. Kyberg and H. Smokler eds. 1964).

32. See section I(A) supra.

33. See section I(B) supra.

34. If Y is the disjunction of all other sufficient conditions, then it follows that $P(Z|AX^{u}Y) = \beta < 1$. This is directly analogous to requirement (2) in the deterministic version.

35. See supra note 34.

36. Suppes calls this a "prima facie" cause. However, in order to avoid confusion with the legal concept of prima facie, the term "naive" is adopted. This was Suppes' first and later discarded name for the idea. See Suppes, supra note 7 at 12, n.1.

37. This is essentially Suppes's definition of a prima facie cause (Id. at 12) with two adjustments. First, the possibility condition, $P(A_{t_o}) > 0$, is eliminated because our ex post method starts from events

that have happened. Second, the negation of A_{t_o} and the environmental conditions (X_{t_o}) are made explicit on the right hand side of the inequality in (2).

38. No claim is being made that the probabilistic relationships that precede and follow exhaust the meaning of causation. Instead, causal connections entail these relationships as one very important set of characteristics.

39. This is a modification of Suppes definition of spurious cause (see Suppes, supra note 7 at 23). (1) In condition (1) we add $<$ as Suppes indicates might be better; (2) In condition (2) we drop the equals sign because it seems reasonable to demand that A_{t_o} increase the explanatory power of A_{t_1}; (3) For our purposes a spurious cause can coexist a genuine cause and the latter need not precede the former.

40. Compare Reichenbach's definition of "causal relevance" (see H. Reichenbach, The Direction of Time 204 (1956)):

> The event A_{t_o} is causally relevant to a later event A_{t_2} if
>
> $$P(A_{t_o}, Z_{t_2}) > P(\bar{A}_{t_o}, Z_{t_2})$$
>
> and there exists no set of events $A^{(1)} \ldots A^{(n)}$ which are earlier than or simultaneous with A_{t_o} such that this set screens off A_{t_o} from Z_{t_2}.
>
> $A^{(2)}$ is said to screen off A_{t_o} from Z_{t_2} if $P(Z_{t_2} | A_{t_o}, A^{(2)}) = P(Z_{t_2} | \bar{A}_{t_o}, A^{(2)})$.

The relationship between Suppes' definition of naive cause and the first part of Reichenbach's definition of causal relevance can be seen easily (suppressing the background or environment term for simplicity):

$$P(A_{t_o}, Z_{t_2}) = P(Z_{t_2} \mid A_{t_o}) \cdot P(A_{t_o}) \text{ (by definition)}$$

$$P(\overline{A}_{t_o}, Z_{t_2}) = P(Z_{t_2} \mid \overline{A}_{t_o}) \cdot P(\overline{A}_{t_o}) \text{ (by definition)}$$

$$P(Z_{t_2} \mid A_{t_o}) \cdot P(A_{t_o}) > P(Z_{t_2} \mid \overline{A}_{t_o}) \cdot P(\overline{A}_{t_o}) \text{ (Reichenbach condition)}$$

$$P(Z_{t_2} \mid A_{t_o}) > P(Z_{t_2} \mid \overline{A}_{t_o}) \cdot P(\overline{A}_{t_o})/P(A_{t_o})$$

The final inequality is the Reichenbach condition in terms easily comparable to Suppes'. When the last term on the right is one the two conditions are the same.

Reichenbach's screening off definition is not as useful as Suppes' spurious cause. The former does not tell us, in probabilistic terms, what screens off what. See W. Salmon, Statistical Explanation and Statistical Relevance 54-55 (1971).

41. See Yeager, Essential Properties of the Medium of Exchange, 21 Kyklos 45, 46 (1968).

42. By "objective" is meant society's best guess or the best-informed probability.

43. See Rosen, supra note 7 at 611.

44. This includes those conditions that are highly probable at the time of A_{t_o}.

45. This is consistent with the thin skull rule in the common law and with Rümelin's use of the hindsight principle in German law.

On the latter, see A. Honoré, Causation and Remoteness of Damage in 11 Int'l Ency. Comp. L. ch. 7, at 51 (1971).

46. This is the prevailing view in German law. Id. at 52.

47. See text accompanying note 26 supra.

48. See, e.g., DeFinetti supra note 31 and L. Savage, Foundations of Statistics (1954).

49. This, of course, is not without its ambiguities. However, in any framework, a court must make judgments about the relative worth of information.

50. See Rizzo and Arnold, supra note 4 at 1407-1410.

51. Id. at 1402-1405.

52. See Restatement (Second) of Torts §435, Comment c and Hart and Honoré, supra note 8 at 160-161.

53. In contrast, the adequate cause theory has the additional requirement that the wrongful conduct be a necessary condition of the harm. See A. Honoré, supra note 45 at 49. This is not a fully probabilistic analysis but confounds probabilistic and deterministic conceptions of causation.

54. This is similar to the increased-risk doctrine, the leading test of compensability under workmen's compensation statutes. See A. Larson, 1 The Law of Workingmen's Compensation, sec. 6.30 (1952).

55. 191 Pa. 345, 43 A. 240 (1899).

56. On the economic efficiency approach, coincidental accidents ought to be included in the scope of liability if the losses are large relative to the value of the activity. It is possible that expected losses involving, say, a bus going over a certain route are so high that the bus ought not to go out at all. Sometimes this result can be achieved only by holding the bus liable regardless of the coincidental nature of the accident (see Shavell, supra note 5, at 479). Neither our approach nor the law makes the outcome in such cases depend on the extent of the losses.

57. See the related matters in section I(A) in Part II infra.

58. See, generally, Honoré, supra note 45 at 83.

59. See Mackie, supra note 9 at 251 for a deterministic example. See also B. Skyrms, Causal Necessity 109 (1980).

60. Honoré, supra note 45 at 70.

61. Id. at 80. See also Hart and Honoré, supra note 8 at 229 and Restatement, supra note 52 at §432, Comment 6.

62. Shavell implicitly recognizes this point. He discusses the coincidental case in which a speeding car will hit cattle crossing the road at point B with a probability .02 and a nonspeeding car will hit them with probability .02 at point A (see Shavell, supra note 5 at 474-475). ". . . [S]peeding is not a cause of expected losses given that the cattle

cross the road at some point" (Id. at 474). The proviso "given that the cattle cross the road" means that the type of accident is being held constant and hence that we are not dealing purely with expected losses.

63. See, e.g., Beale, The Proximate Consequences of an Act, 33 Harv. L. Rev. 633, 644-645 (1920), and McLaughlin, Proximate Cause, 39 Harv. L. Rev. 149, 160-169 (1925)

64. More detailed discussion of intervening events can be found in section IV(A) infra.

65. Suppes' definition of a direct cause (see Suppes, supra note 7 at 28) implies that unless A_{t_o} can be totally factored out in condition (iii) (i.e., unless $P(Z_{t_2} \mid A_{t_o}, I_{t_1}, X_{t_o}) = P(Z_{t_2} \mid I_{t_1}, X_{t_o})$) then A_{t_o} is a direct cause. This requires the causal significance of the intervention to be "complete" before it can negate the causal connection between A_{t_o} and Z_{t_2}. Since the law usually requires a less significant causal contribution from I_{t_1} the above definition was thought more appropriate in this context.

66. See Salmon, supra note 40 at 10-11, 56-57.

67. The standard environmental variable, X_{t_o}, will be suppressed from now on for simplicity.

68. See McLaughlin, supra note 63 at 181.

69. See section III(A) of part II infra.

70. In re Polemis & Furness, Withy, & Co., [1921] 3 K.B. 560.

71. See text accompanying note 11 supra.

72. The Polemis viewpoint is law in most American jurisdictions (see C. Gregory, H. Kalven, Jr. and R. Epstein, Cases and Materials on Torts 279-280 (3rd ed. 1977) and Christianson v. Chicago, St. P., M. & O. Ry., 67 Minn. 94, 69 N.W. 640 (1896)). Overseas Tankship (U.K.) ltd. v. Morts Dock & Eng'r Co., [1961] A.C. 388 (P.C. Aust.) (Wagon Mound I) has apparently had little, if any, influence in changing the courts' adherence to the Polemis rule. For an explicit American rejection of Wagon Mound I see Petition of Kinsman Transit Company 338 F.2d 708 (2d Cir. 1964). For an analysis of the unimportance of Wagon Mound I to subsequent cases in England and Scotland, see Rizzo and Arnold, supra note 4 at 1405, n. 33.

73. Recall that the X_{t_o} terms have been suppressed for simplicity.

74. Suppose, on the other hand, that the collected benzine vapor was either due to the plaintiff's negligence or was a nonnnegligent accident. Then the independent harm-producing quality of the benzine would decrease the likelihood that the defendant would be found liable (or it would decrease his share in an apportionment scheme). Assume, for simplicity, that the probability of the ship burning absent both the dropped plank and the vapor is zero (α). If the probability given only the preexisting vapor is β and the probability given both the vapor and fall of the plank is γ, then the defendants are liable if and only if $(\gamma-\beta) > (\beta-\alpha)$. In other words, they are liable if the

marginal product of the plank-dropping exceeds that of the benzine vapor. Where $\alpha = o$, this reduces to the more familiar $(\gamma-\beta) > \beta$.

75. See another example in Epstein, A Theory of Strict Liability, 2 J. Legal Stud. 151, 184 (1973).

76. 118 Mass. 251 (1875).

77. It is incorrect to argue that since the marginal product of A at t_o is lower than the marginal product of M at t_1 that M_{t_1} is the true cause. To see this, consider that since $(\beta-\alpha) > \alpha$, M_{t_1} is itself causally dependent on a prior event A_{t_o} while, ex hypothesi, A_{t_o} is not so dependent.

78. A more recent case of this type is Lynch v. Fisher (34 So. 2d 513 (1947)). The wrongful conduct of the defendant caused an automobile collision in which an injured individual became trapped in his car. The plaintiff, a bystander, attempted to help him. Instead, the trapped man, after being handed a gun that the plaintiff found lying in the area, shot plaintiff. It was established that the shock of the accident caused the third party to do this. Plaintiff recovered from defendant. This is another example of the intermediate event (M_{t_1}), the shooting, being dependent on the original wrong (A_{t_o}). Since the marginal product of A_{t_o} in producing M_{t_1} is greater than $\overline{A}_{t_o}$, the marginal product of A_{t_o} in producing the ultimate injury (Z_{t_2}) exceeds $\overline{A}_{t_o}$. Hence A_{t_o} is a proximate cause of Z_{t_2}.

79. See, e.g., Beale, supra note 63 at 644-645; Carpenter, Workable Rules for Determining Proximate Cause (Part III), 20 Cal. L.

Rev. 471, 474 (1932); McLaughlin, supra note 63 at 160-167; Myers, Causation and Common Sense, 5 Miami L. Q. 238, 242-243 (1951) and all of the cases cited in the above.

80. Beale (supra note 63 at 645) says that the direct result of a passive cause, an omission, is not always proximate. However, he presents only two cases and one of them concludes that the omission was proximately related to the harm. See Terre Haute & I.R.R. v. Buck, 96 Ind. 346 (1884).

81. Myers, supra note 79 at 245.

82. See, e.g., Engle v. Director General of Railroads, 78 Ind. 547, 133 N.E. 138 (1921). This is cited by Myers (Id. at 245, n. 40) in support of his claim that there are exceptions.

83. The second of the two cases cited by Myers (Id. at 245-246, n. 40) in support of the exception claim is not a direct harm at all (as Myers himself admits, relying instead on dictum in the case). See Cone v. Inter-County Tel. & Tel. Co., 40 So. 2d 148 (Fla. 1949).

84. 248 N.Y. 339, 162 N.E. 99 (1928).

85. See section IV(B)(2) infra.

86. See Rizzo and Arnold, supra note 4 at 1407-1408.

87. See section I(B) of Part II supra.

88. But see Rizzo and Arnold, supra note 4 at 1417-1418.

89. Compare this statement with conditions (ii) and (iii) of superseding events. See section III(A) of Part II supra.

90. For an emphasis on the importance of foreseeability of subsequent conduct see W. Prosser, The Law of Torts §44, at 274 (4th ed. 1971). But see Rizzo and Arnold, supra note 4 at 1420-1421 for an explanation of the role of various kinds of conduct in completing or interrupting the chain of causation without regard to the foreseeability or probability of intervention. Although most of the cases can be explained in this way, there are exceptions, see Hart and Honoré, supra note 8 at 179-184.

91. Since in actual cases precise probability measures are rarely possible, the law treats superseding and concurrent causation in an approximate fashion. The greater the deviation of $(\beta-\alpha)$ from α the more likely the second cause will be found to be a superseding cause; the smaller that deviation the more likely the two causes will be found to concur.

92. But see the apportionment mechanism in Rizzo and Arnold, supra note 4 at 1419.

93. Technically, when the causes are simultaneous neither can supersede the other since, by definition, the superseding cause is later in time than the other. See Restatement, supra note 52 at §440. The terminology here is usually that of "material" and "substantial" factor. See section IV(B)(3) infra.

94. See text accompanying note 52 supra.

95. See, however, McLaughlin, supra note 63 at 182-183 who defends the minority position that unforeseeable here means improbable from the perspective of the reasonable man.

96. Restatement, supra note 52 at §435, Comment c (emphasis added).

97. Id. at §435, Comment d. See also W. Prosser, supra note 90 §44, at 272.

98. "The court's judgment . . . is made . . . after the event with the full knowledge of all that has happened. This includes those surroundings of which at the time the actor knew nothing but which the course of events discloses to the court." Restatement, supra note 52 at §435, Comment d.

99. See section V(B)(1) infra.

100. 97 Ohio St. 262, 119 N.E. 733 (1918).

101. This is another way to analyze this case that is perfectly consistent with the underlying theory and gives the same basic result. Let α be the probability of the goods being destroyed by a flood at t_o given only the standard environment (i.e., nonnegligent shipping and other environmental conditions) and β that probability at t_o given negligent shipping. Ex ante, it is clear that $\alpha \cong \beta$ and hence the marginal product of wrongful conduct is approximately zero. On this analysis Toledo is a coincidence case.

102. 248 N.Y. 339, 162 N.E. 99 (1928).

103. Id. at 340, id. at 99 (emphasis added).

104. The reasons given for the decision in Palsgraf are obviously inconsistent with the claim that foreseeability has no role in direct harms since the logic of "unforeseeable plaintiffs" is applicable to direct harms as well. However, the doctrine of Palsgraf has not had much influence on the courts (as opposed to the legal scholars). See C. Gregory, H. Kalven, Jr. and R. Epstein, supra note 72 at 260 and Hart and Honoré, supra note 8 at 246.

105. 146 Minn. 430, 179 N.W. 45 (1920).

106. See Restatement, supra note 52 at §§431, 432 and 433.

107. See Hart and Honoré, supra note 8 at 218.

108. The court applies the substantial factor test even when both fires would, taken individually, be sufficient to cause the harm (see id.). This may appear to be in conflict with the probabilistic analysis. Actually, it is not. If probability is interpreted as the relative frequency, in a large number of similar cases, of one fire alone resulting in the harm, then both fires could have been sufficient in this case and yet only one is substantial.

109. Shavell, supra note 5 at 481.

110. This assumes that the first best solution is to engage in the activity. If not, then imposing liability where the marginal product

of care is low but where losses are high will properly discourage people from engaging in the activity.

111. ". . . [W]e may formulate a hypothesis which covers most of the cases of liability without fault . . . [S]ome types of conduct create such risks of harms to others, even when the conduct itself is careful, that, though the creation of the risks is not itself unreasonable because of the social utility of the conduct, nevertheless it is unreasonable for the person injured to bear his own loss." F. Harper, Liability Without Fault And Proximate Cause, 30 Michigan L. Rev. 1001, 1005-1006 (emphasis added) (1932). "Thus, the duty concept is inappropriate to characterize the strict liability cases" (Id. at 1014). "It is much more intelligible to state that the defendant has engaged in a certain type of conduct which is a sufficient basis of liability . . ." (Id. at 1015).

112. Shavell, supra note 5 at 448.

113. "Most writers, however, hold that it is the tortfeasor's fault which must be shown to have caused the harm." Honoré, supra note 45 at 18.

114. See Hart and Honoré, supra note 8 at 111-114.

115. Shavell is well-aware that the law is not consistent with these implications of his efficiency theory (see Shavell, supra note 5 at 489). The reason that the law under negligence looks as it "should" under strict liability (i.e., liability is based on the marginal product

of taking care) is because courts tend to set due care standards incorrectly and hence many defendants are found negligent who shouldn't be. This imparts an element of strict liability to the system (Id.). There are several problems with this argument. (1) The courts must set the due care level too high, otherwise no one will have to pay for accidents that occur despite optimal precautionary activity. If they are set too low, however, there is no element of strict liability. Hence Shavell must explain why there is a systematic upward bias in the court's error mechanism (or at least show that, in fact, there is such a bias). (2) If this bias exists, we should observer that under strict liability the marginal product of taking care is relatively more important to the determination of proximate cause than under negligence. Certainly, the reverse is more accurate a description of the existing law.

116. See id. at 484-485.

117. By "expected" is meant discounted by the probability of occurrence.

118. See R. Posner, supra note 1 at 130 and Posner, A Theory of Negligence, 1 J. Legal Stud. 29, 42 (1972).

119. The former do not negative causal connection but the latter may. See Hart and Honoré, supra note 8 at 160-164.

120. Why people should underestimate rather than overestimate these events is not obvious on theoretical grounds. Furthermore, the cognitive biases uncovered by experimental psychology do not all

tend to produce underestimation. Sometimes these biases may generate overestimation as when people find it easy to imagine instances of airplane disasters or other low probability events because of media publicity ("the ease of constructing instances does not always reflect their actual frequency"). Furthermore, people are often "anchored" to an excessively high (or low) prior probability estimate (insufficient adjustment). Thus, "people tend to overestimate the probability of conjunctive events and to underestimate the probability of disjunctive events." See A. Tversky and D. Kahneman, Judgment Under Uncertainty: Heuristics and Biases, 185 Science 1124, 1127, 1129 (1974).

121. See Shavell, supra note 5 at 490-491.

122. Shavell's analysis is not correct for the general case in which care can be varied in continuous gradations. The level of due care is set by, let us suppose, the economically omniscient court. Even if (due to underestimation of a particular state of the world) there is not sufficient incentive to get defendant to adhere to due care, it does not follow that he will exercise no care at all. Since he will be liable for plaintiff's injuries he will want to engage in some amount of avoidance. Holding defendant liable even in underestimated states will contribute to that. Suppose, for example, that the true expected net marginal product of another unit of avoidance is $5 but due to underestimation the defendant sees it as only $2. Suppose, further, that the true expected administrative costs are $3. It follows that defendant ought to be held liable in this state of the world because we ought to compare the administrative costs with the

true expected marginal product of avoidance. It is true that defendant will not push avoidance to the optimal level but eliminating the underestimated state from the scope of liability only makes matters worse.