

Counterfactual Causation

Hillel J. Bavli*

ABSTRACT

Causation is commonly defined using the counterfactual model, and the “but-for” standard in particular. It asks whether the harm suffered by the plaintiff would have occurred in the absence of the defendant’s act. It is commonly believed, however, that the counterfactual model fails in cases involving multiple sufficient causes—that is, cases in which two or more forces contribute to an outcome where each force alone would suffice to produce the same outcome. This paradox has, over time, pushed causation standards into a state of ambiguity and disarray as courts have attempted to retain the counterfactual model as the appropriate framework for causation while abandoning it in multiple-sufficient-cause situations to attain the sought-after outcome.

In this article, I argue that, contrary to common understanding, the counterfactual model does not fail in multiple-sufficient-cause situations. In particular, I propose the adoption of a framework for cause and effect in statistics and the sciences called the “potential outcomes framework,” and I apply it to explain and address the apparent paradox of multiple sufficient causes. I then extend my analysis to show a broad range of implications for fields such as torts, criminal law, contracts, constitutional law, and employment discrimination. Beyond demonstrating important consequences for standards of causation in various substantive areas of the law, I show how my analysis affects our understanding and treatment of timely issues, such as the judicial interpretation of causal language in criminal statutes and the permissibility of “mixed-motive” cases under Title VII and other federal discrimination laws.

* Assistant Professor of Law, SMU Dedman School of Law; Fellow, Harvard Institute for Quantitative Social Science. The author wishes to thank Gregory Crespi, Michael Green, Christopher Robertson, Edward Cheng, Iavor Bojinov, Barry Goldstein, David Rosenberg, Daniel Markovits, Louis Kaplow, Donald Rubin, Jameson Quinn, J. Shahar Dillbary, and Guillaume Basse for their helpful comments, and Sophie Franks for her research assistance. The author is grateful for the generous financial support provided by SMU Dedman School of Law and the WWB Law Professor’s Fund.

I. INTRODUCTION

Factual causation is the element of a legal claim that requires a plaintiff to link the defendant's conduct to the plaintiff's injury—to show that the former did not merely coincide with the latter but rather produced it. It is intended to reflect “natural” or “actual” cause and effect, a “scientific” causal connection between conduct and injury.¹ It aims to capture “our common understanding of causation” and “deep-seated intuitions about causation and fairness in attributing responsibility.”² Developing an appropriate standard, or even definition, of causation is a perennial problem across many areas of the law. Strangely, however, although courts frequently look to statistics and the sciences for *evidence* of causation—sometimes even requiring such evidence to prove causation—the law has rarely looked to these fields for guidance in developing a suitable *definition* of causation. This is in spite of a well-studied and robust field of science devoted to questions of cause and effect.³

In this article, I propose the adoption of a pervasive framework for conceptualizing and answering questions of cause and effect in statistics and the sciences to inform the meaning of causation in law. It is called the potential outcomes framework, or the “Rubin Causal Model.” In particular, I aim to establish the applicability of the framework and to apply it to address a broad class of complex and longstanding problems.

Courts and scholars frequently misunderstand the notion of actual or scientific cause and effect. In many cases this is without consequence, and a simple understanding suffices. But, in one common circumstance, this misunderstanding has wreaked havoc across a wide range of legal fields: situations involving “multiple sufficient causes.” In these cases, in which two

1. See DAN B. DOBBS ET AL., HORNBOOK ON TORTS 314 (2d ed. 2016) (“These causal issues raise questions of fact in the scientific sense.”); W. PAGE KEETON ET AL., PROSSER AND KEETON ON THE LAW OF TORTS § 41, at 264–65 (5th ed. 1984) (“This question of ‘fact’ ordinarily is one upon which all the learning, literature and lore of the law are largely lost. It is a matter upon which lay opinion is quite as competent as that of the most experienced court.”); Richard W. Wright, *The NESS Account of Natural Causation: A Response to Criticisms*, in PERSPECTIVES ON CAUSATION 285, 285 (Richard Goldberg ed., 2011) (“natural (scientific, ‘actual’, ‘factual’) causation”); see also *infra* note 79 and accompanying text. Distinguish “scientific,” and even “statistical,” cause and effect from “probabilistic” notions of cause and effect. By “scientific” or “statistical” cause and effect, I refer simply to the way in which scientists or statisticians conceptualize cause and effect, whether “actual” or probabilistic, “ex post” or “ex ante.” See *infra* section III.A.

2. RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL & EMOTIONAL HARM § 27 cmt. c (AM. LAW INST. 2010). Distinguish factual causation from proximate causation (or “scope of liability”), which restricts the field of liability to certain harms. *Id.* § 29.

3. Throughout this article, I refer to the generic causal concept as “cause and effect,” the legal causal concept as “causation,” and the statistical or scientific causal concept as “causality” or “causal inference.”

or more forces contribute to produce an outcome (such as a plaintiff's injury), where each force alone would suffice to produce the same outcome, a more nuanced understanding of cause and effect is needed. Indeed, multiple-sufficient-cause situations arise frequently and, as I will show, have important implications for a wide range of legal fields, including torts, criminal law, employment discrimination, constitutional law, and contracts, among others. The potential outcomes framework is applicable to these situations and instructive for solving the important and enduring problems to which they give rise.

The most well-accepted and pervasive test of factual causation in the law is the "but-for," or *sine qua non*, test. It uses a counterfactual model of causation: it asks whether an outcome would have occurred absent the alleged conduct. The but-for inquiry seeks to determine whether the defendant's act was a necessary condition of the outcome. This standard is generally simple to apply and, in most cases, leads to clear and uncontroversial results. For example, in a case involving a driver who negligently ran a red light and struck a pedestrian, it is straightforward to infer that the driver's negligent act is the but-for cause of the pedestrian's injuries.

In cases involving multiple sufficient causes, however, the but-for test is said to fail.⁴ It is often said, and courts have generally held, that in these situations "some other test is needed."⁵ There are two categories of cases involving multiple sufficient causes (also referred to as "overdetermined causation" cases⁶): concurrent multiple sufficient causes and successive multiple sufficient causes.⁷ The "concurrent" category involves cases in which two or more forces occur concurrently to produce an injury—for example, two separate fires merge to destroy a plaintiff's lodge, where each fire would have been sufficient to destroy it. Under the traditional but-for standard, the originator of each fire could assert that she is not liable for the damage because, even in the absence of her fire, the other fire would have destroyed the plaintiff's lodge. The "successive" category involves cases in which two or more forces occur successively, with an earlier force "preempting" the effect of a second force. This would occur, for example, if

4. *Major v. R.J. Reynolds Tobacco Co.*, 222 Cal. Rptr. 3d 563, 581 (Ct. App. 2017) (stating that "[t]he law is . . . clear" that the but-for test leads to the incorrect result in cases involving multiple sufficient causes); *KEETON ET AL.*, *supra* note 1, § 41, at 266 ("[T]here is one type of situation in which [the but-for test] fails. If two causes concur to bring about an event, and either one of them, operating alone, would have been sufficient to cause the identical result, some other test is needed.").

5. *KEETON ET AL.*, *supra* note 1, § 41, at 266.

6. Richard W. Wright, *Causation in Tort Law*, 73 CALIF. L. REV. 1735, 1740 (1985).

7. These are sometimes referred to as "duplicative" and "preemptive" causation scenarios. *See id.* at 1775.

one of the above fires alone destroyed the plaintiff's lodge before the second fire arrived, but where the second fire would have been sufficient to destroy the lodge had the first fire not already destroyed it.⁸ Under the traditional but-for standard, the originator of the first fire could, as above, claim that she is not liable for the damage because, even in the absence of her fire, the second fire would have destroyed the lodge.

Courts and scholars have arrived at a near-universal consensus that but-for causation is incapable of appropriately handling cases involving multiple sufficient causes. As a leading torts treatise states, “[t]he but-for test in such cases leads to a result that is almost always condemned as violating both an intuitive sense of causation and good legal policy.”⁹ Specifically, it is said that, in either duplicative or preemptive causation situations, “contrary to the laws of nature, common sense, and the decisions of the courts, neither fire [in the two-fire problem] would be treated as a cause under the strong necessity criterion [the principle feature of but-for causation], since neither was strongly necessary given the existence of the other.”¹⁰

Courts, therefore, have commonly rejected the but-for standard in cases involving multiple sufficient causes, holding that it “should not be used when two ‘causes concur to bring about an event and either one of them operating alone could have been sufficient to cause the result.’”¹¹ Instead, courts generally apply a vague “substantial-factor” test that relies on the intuition of the factfinder to make a causal determination. Moreover, the purported failure of the but-for test has prompted calls to abandon the counterfactual model altogether as a theory of factual causation.¹² It has been argued, for example, that “[u]nless appearances are deceiving, [these] cases by themselves show that the counterfactual theory cannot be a theory of causation.”¹³

It is frequently thought that multiple-sufficient-cause situations arise primarily in torts cases. But these situations, and the purported failure of the but-for test, arise frequently in a wide range of legal contexts. For example, they arise in contracts cases when multiple contractors fail to timely perform under a contract, where each breach alone would suffice to cause the

8. See DOBBS ET AL., *supra* note 1, at 321–22; KEETON ET AL., *supra* note 1, § 41, at 266–67; Wright, *supra* note 6, at 1775–76.

9. DOBBS ET AL., *supra* note 1, at 321–22.

10. Richard W. Wright & Ingeborg Puppe, *Causation: Linguistic, Philosophical, Legal and Economic*, 91 CHI.-KENT L. REV. 461, 474 (2016).

11. *Mitchell v. Gonzales*, 819 P.2d 872, 876 (Cal. 1991) (quoting *Vecchione v. Carlin*, 168 Cal. Rptr. 571, 576 (Ct. App. 1980)).

12. See *infra* note 24 and accompanying text.

13. MICHAEL S. MOORE, CAUSATION AND RESPONSIBILITY: AN ESSAY IN LAW, MORALS, AND METAPHYSICS 411 (2009); see also Note, *Rethinking Actual Causation in Tort Law*, 130 HARV. L. REV. 2163, 2165 (2017).

plaintiff's damages. Perhaps most significantly—in light of the current state of confusion and controversy in this area of the law—“mixed-motive” employment discrimination cases can and should be understood as a special type of multiple-sufficient-cause situation. These cases involve an adverse employment action (e.g., firing an employee) that is motivated by both legitimate purposes (e.g., tardiness or rudeness to customers) and illegitimate purposes (e.g., race, gender, or age), where each purpose alone would suffice to produce the adverse action. Congress, the Supreme Court, and lower courts have grappled extensively to establish appropriate standards of causation in these circumstances, with unfortunate results—results that have led to vague and contradictory standards under Title VII, the Americans with Disabilities Act, and other discrimination statutes.

In this article, I demonstrate the applicability of the potential outcomes framework for considering factual-causation problems in the law. I then apply it to explain the purported failure of the but-for model of causation in multiple-sufficient-cause situations. In particular, I show that but-for causation is only a specific measure of interest (in statistics terminology, a specific “estimand”) within the broader counterfactual model applicable to factual-causation inquiries. It is incorrect to conclude that counterfactual causation fails in multiple-sufficient-cause situations; and, it is oversimplified to conclude that the but-for standard in particular fails in these situations. I apply the potential outcomes framework—a predominant framework in the sciences and a counterfactual model of cause and effect—to establish a more nuanced understanding of causation in law. Then, I argue that a refined conception of causation based on the potential outcomes framework carries a wide range of implications for causation problems across various areas of the law. In particular, I argue the following:

1. My findings discredit efforts to abandon the counterfactual model of causation on the basis of its purported failure in multiple-sufficient-cause situations.

2. Courts should reject the vague and intuition-based substantial-factor test of causation—a test that is prevalent in many areas of the law, notwithstanding its inadequacies—and, in cases involving multiple sufficient causes, should apply a counterfactual model that follows the potential outcomes framework. In practice, this approach can be simplified and applied by using the causal-set theory of the “NESS test” (defined below) and the Restatement (Third) of Torts, and, therefore, can be understood as providing a scientific foundation for these approaches.

3. The potential outcomes framework informs judicial interpretation of the “ordinary meaning” of causal statutory language such as “because of,” “results from,” or “based on,” which courts frequently apply when a statute

does not specify a term's meaning. My analysis could apply to resolve apparent conflicts between the ordinary meaning of a statute's causal language and Congress's intent to capture factors in multiple-sufficient-cause situations as "causes."

4. Finally, my analysis implies that, in mixed-motive employment discrimination cases, a motivating factor could be understood as a cause of an adverse employment action even when other sufficient motivating factors exist; and, as such, it could address contradictory, vague, and potentially inappropriate standards governing causation in these cases.

I begin in Part II by introducing the problem of multiple sufficient causes. I describe why multiple-sufficient-cause situations and the application of the but-for standard in these situations are viewed as problematic, and how courts and scholars have addressed the issue. In Part III, I introduce the potential outcomes framework and its implications for factual causation. In Part IV, I apply a counterfactual model of causation that follows the potential outcomes framework to analyze concurrent and successive multiple-sufficient-cause problems. I show that, although an unrefined application of the counterfactual model fails in multiple-sufficient-cause situations, a more nuanced understanding of this model yields results that are consistent with common sense, policy, and court decisions. Because the problem of multiple sufficient causes has been litigated and discussed extensively in the torts context, and because other fields of law frequently refer to tort law when addressing the issue, I couch my argument in Parts II–IV in the torts setting. Then, in Part V, I discuss implications for a wide range of legal contexts, including torts, criminal law, contracts, and employment discrimination. In Part VI, I conclude.

II. THE BUT-FOR STANDARD IN MULTIPLE-SUFFICIENT-CAUSE SITUATIONS

Factual causation is widely understood as fundamental to legal liability, and to tort liability in particular.¹⁴ It is intended to reflect cause and effect in

14. See DOBBS ET AL., *supra* note 1, at 317; MARC A. FRANKLIN ET AL., TORT LAW AND ALTERNATIVES: CASES AND MATERIALS 337–95 (10th ed. 2016); David W. Robertson, *Causation in the Restatement (Third) of Torts: Three Arguable Mistakes*, 44 WAKE FOREST L. REV. 1007, 1008 (2009) (“[T]he cause-in-fact requirement is the ‘linchpin’ of the corrective-justice theory [of tort law]. Indeed, it has long been regarded as a truism that ‘a defendant should never be held liable to a plaintiff for a loss where it appears that his wrong did not contribute to it, and no policy or moral consideration can be strong enough to warrant the imposition of liability in such [a] case.’”) (quoting Charles E. Carpenter, *Concurrent Causation*, 83 U. PA. L. REV. 941, 947 (1935)); John D. Rue, Note, *Returning to the Roots of the Bramble Bush: The “But for” Test*

the common, or “scientific,” sense, which, in turn, usually refers to counterfactual, or (in particular) but-for, cause and effect.¹⁵ As the Supreme Court has stated:

A thing “results” when it “[a]rise[s] as an effect, issue, or outcome *from* some action, process or design.” “Results from” imposes, in other words, a requirement of actual causality. “In the usual course,” this requires proof “‘that the harm would not have occurred’ in the absence of—that is, but for—the defendant’s conduct.”¹⁶

The but-for condition is central to the concept of causation.¹⁷ However, applying the but-for standard to cases involving multiple sufficient causes gives rise to a paradox: despite the seemingly sound and well-accepted reasoning of the counterfactual model, and the but-for standard in particular, applying this reasoning leads to a seemingly illogical conclusion.

Consider the following situation:

Rosaria and Vincenzo were independently camping in a heavily forested campground. Each one had a campfire, and each negligently failed to ensure that the fire was extinguished upon retiring for the night. Due to unusually dry forest conditions and a stiff wind, both campfires escaped their sites and began a forest fire. The two fires, burning out of control, joined together and engulfed Centurion Company’s hunting lodge, destroying it. Either fire alone would have destroyed the lodge.¹⁸

Regains Primacy in Causal Analysis in the American Law Institute’s Proposed Restatement (Third) of Torts, 71 *FORDHAM L. REV.* 2679, 2679–80 (2003) (“An empirically ascertainable connection between a defendant’s tortious act and a plaintiff’s injury (‘factual cause’) has long been considered a threshold question of liability.”) (citing WILLIAM L. PROSSER, *TORTS* § 41 (4th ed. 1971)). Note that law and economics scholars have often *not* viewed ex post cause and effect as necessary to promote the efficiency aims of tort law (and have sometimes viewed it as detrimental to such aims); but they at least recognize the importance of *some* causal link between misconduct and harm. *See infra* note 48.

15. *See supra* note 1; *infra* note 79.

16. *Burrage v. United States*, 571 U.S. 204, 210–11, 213–14 (2014) (citations omitted) (citing civil and criminal cases); *see also Paroline v. United States*, 572 U.S. 434, 452 (2014) (“These alternative causal tests [to but-for causation] are a kind of legal fiction or construct.”).

17. *See FRANKLIN ET AL.*, *supra* note 14, at 337 (“The core of causation is that if X had not occurred, Y would not have occurred. This requirement is frequently referred to as ‘but for’ or *sine qua non*.”).

18. *RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL & EMOTIONAL HARM* § 27 cmt. a, illus. 1 (AM. LAW INST. 2010). *See generally DOBBS ET AL.*, *supra* note 1, at 321 (citing *Landers v. E. Tex. Salt Water Disposal Co.*, 248 S.W. 2d 731 (Tex. 1952), a case in which two defendants were alleged to have independently caused salt water and oil to spill into plaintiff’s lake, killing plaintiff’s fish); Wex S. Malone, *Ruminations on Cause-in-Fact*, 9 *STAN. L. REV.* 60, 89 (1956)

Our common intuition leads us to identify an event as a cause of an outcome if the outcome would occur in the presence of, but not in the absence of, the event—that is, if the outcome would not occur but for the event. This description characterizes the but-for conception of causation. Applying this standard to the example above, however, each camper could avoid responsibility, because the other camper’s fire would have alone caused the same result—the destruction of Centurion Company’s hunting lodge. It is practically a consensus that the but-for standard leads us astray in this type of situation, that it fails relative to our common sense of cause and effect in cases involving multiple sufficient causes.¹⁹ “Our senses have told us that [the defendant] *did* participate. . . . In the language of the layman, the defendant’s fire ‘had something to do with’ the burning of plaintiff’s property.”²⁰ It has been said, therefore, that “the but-for test denies the existence of cause in fact while everything in human experience and intuition cries out that cause in fact was present.”²¹ These results are said to be “contrary to the laws of nature, common sense, and the decisions of the courts.”²²

This quandary has generally led courts to hold simply that “the ‘but for’ test . . . should not be used when two ‘causes concur to bring about an event and either one of them operating alone could have been sufficient to cause the result.’”²³ Some courts and scholars have gone even further, arguing that

(“Two noisy motorcycles simultaneously pass the horse on which plaintiff is riding, frightening the animal and causing it to run away and injure [the] plaintiff. . . . Similarly, a fire started through the negligence of a railroad may merge with a fire of undetermined origin and the two together destroy plaintiff’s property.”) (citing *Corey v. Havener*, 65 N.E. 69 (Mass. 1902); *Anderson v. Minneapolis, St. P. & S. Ste. M. Ry.*, 179 N.W. 45 (Minn. 1920)).

19. See RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL & EMOTIONAL HARM § 27 cmt. c (AM. LAW INST. 2010) (“[W]hile the but-for standard provided in § 26 is a helpful method for identifying causes, it is not the exclusive means for determining a factual cause. Multiple sufficient causes are also factual causes because we recognize them as such in our common understanding of causation, even if the but-for standard does not. Thus, the standard for causation in this Section comports with deep-seated intuitions about causation and fairness in attributing responsibility.”).

20. Malone, *supra* note 18, at 89 (commenting on the situation in which a railroad negligently started a fire, which then merged with another fire and destroyed plaintiff’s property).

21. David W. Robertson, *The Common Sense of Cause in Fact*, 75 TEX. L. REV. 1765, 1777 (1997).

22. Wright & Puppe, *supra* note 10, at 474; see also *Major v. R.J. Reynolds Tobacco Co.*, 222 Cal. Rptr. 3d 563, 582 (Ct. App. 2017) (rejecting the but-for test, explaining that absent an alternative to the but-for test, “each of three equally liable [cigarette producer] tortfeasors can escape liability on the basis that they are neither but-for causes nor concurrent independent causes—a wholly unjust result”).

23. *Mitchell v. Gonzales*, 819 P.2d 872, 876 (Cal. 1991) (quoting *Vecchione v. Carlin*, 168 Cal. Rptr. 571, 576 (Ct. App. 1980)).

the failure of the but-for standard as applied to multiple-sufficient-cause situations invalidates it altogether as a model of causation.²⁴

Thus, at least in cases involving multiple sufficient causes, courts have replaced the but-for test with alternative standards, and predominantly the substantial-factor test.²⁵ The substantial-factor test “was suggested as the proper test for cause-in-fact as early as 1911”²⁶ and is used in the first and second Restatements.²⁷ “Its primary function was to permit the factfinder to decide that factual cause existed when there were multiple sufficient causes—each of two separate causal chains sufficient to bring about the plaintiff’s harm, thereby rendering neither a but-for cause.”²⁸ But, as indicated in the Restatement Third, the substantial-factor test “has proved confusing and been

24. See MOORE, *supra* note 13, at 411; Ben Gifford, *State v. Brelo and the Problem of Actual Causation*, 44 AM. J. CRIM. L. 157, 186 (2017) (“As this essay illustrates, cases of concurrent sufficient causation challenge the standard picture of actual causation as but-for causation. This challenge must be resolved either by defending the standard picture—and thereby denying causal status to concurrent sufficient causes—or by modifying or replacing the standard picture. As this essay argues, the standard picture is likely due for a replacement.”); Florence G’sell, *Causation, Counterfactuals and Probabilities in Philosophy and Legal Thinking*, 91 CHI.-KENT L. REV. 503, 512 (2016) (“The test produces counterintuitive results, especially in cases of causal overdetermination. . . . These problems have led some commentators to suggest the abandonment of the but for test.”); Note, *Rethinking Actual Causation in Tort Law*, *supra* note 13, at 2165; see also Rutherford v. Owens-Illinois, Inc., 941 P.2d 1203, 1214 (Cal. 1997) (stating that “California has definitively adopted the substantial factor test of the Restatement Second of Torts,” which “subsumes the ‘but for’ test while reaching beyond it to satisfactorily address other situations, such as those involving independent or concurrent causes in fact”); RESTATEMENT (SECOND) OF TORTS § 431 (AM. LAW INST. 1965) (“The actor’s negligent conduct is a legal cause of harm to another if (a) his conduct is a substantial factor in bringing about the harm”). See generally Wright, *supra* note 6; Wright & Puppe, *supra* note 10.

25. See *Burrage v. United States*, 571 U.S. 204, 215–16 (2014) (“One prominent authority on tort law asserts that ‘a broader rule . . . has found general acceptance: The defendant’s conduct is a cause of the event if it was a material element and a substantial factor in bringing it about.’”) (quoting KEETON ET AL., *supra* note 1, § 41, at 267); see also Rue, *supra* note 14, at 2681 (“[T]he use of the ‘substantial factor’ test has mushroomed, and functions as a part of the causation analysis conducted by courts in ‘virtually every jurisdiction.’”) (quoting Ralph Nader, *The Corporate Drive to Restrict Their Victims’ Rights*, 22 GONZ. L. REV. 15, 16 n.8 (1986)); Wright & Puppe, *supra* note 10, at 480 n.80 (citing cases that have used distinct but similar language, such as “material contribution”).

26. BARRY A. LINDAHL, MODERN TORT LAW: LIABILITY AND LITIGATION § 4:6 (2d ed. 2018); see Jeremiah Smith, *Legal Cause in Actions of Tort*, 25 HARV. L. REV. 103 (1911).

27. See RESTATEMENT (SECOND) OF TORTS §§ 431–32 (AM. LAW INST. 1965); RESTATEMENT OF TORTS §§ 431–32 (AM. LAW INST. 1934).

28. RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL AND EMOTIONAL HARM § 26 cmt. j (AM. LAW INST. 2010); see also KEETON ET AL., *supra* note 1, § 41, at 268 (“The substantial-factor rule was developed primarily for cases in which application of the but-for rule would allow each defendant to escape responsibility because the conduct of one or more others would have been sufficient to produce the same result.”).

misused.”²⁹ Therefore, the Restatement Third rejects the substantial-factor test in favor of the following formulation: “If multiple acts occur, each of which under § 26 [Factual Cause] alone would have been a factual cause of the physical harm at the same time in the absence of the other act(s), each act is regarded as a factual cause of the harm.”³⁰ The Restatement Third adopts a form of the “causal-set” approach of the NESS (necessary element of a sufficient set) test,³¹ under which “a condition contributed to some consequence if and only if it was necessary for the sufficiency of a set of existing antecedent conditions that was sufficient for the occurrence of the

29. RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL AND EMOTIONAL HARM § 26 cmt. j (AM. LAW INST. 2010).

30. *Id.* § 27. See generally KEETON ET AL., *supra* note 1, § 41, at 268 (“It is possible—and more helpful it would seem—to apply an alternative formulation that addresses directly the need for declining to follow the but-for rule in this context. The alternative formulation is this: When the conduct of two or more actors is so related to an event that their combined conduct, viewed as a whole, is a but-for cause of the event, and application of the but-for rule to them individually would absolve all of them, the conduct of each is a cause in fact of the event.”).

31. See RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL AND EMOTIONAL HARM § 26 cmt. c, § 27 cmts. a, f, i (AM. LAW INST. 2010); see also David A. Fischer, *Insufficient Causes*, 94 KY. L.J. 277, 277 (2005); Richard W. Wright, *The New Old Efficiency Theories of Causation and Liability*, 7 J. TORT L. 65, 70 (2014).

consequence.”³² The NESS test has gained momentum,³³ but courts predominantly follow the substantial-factor test endorsed by earlier Restatements.³⁴

The substantial-factor test has been widely criticized. For example, Joseph Sanders and Michael Green have written,

The ambiguity surrounding the substantial-factor test leads to inconsistent results, at least across jurisdictions. More importantly, the test gives no clear guidance to the factfinder about how one should approach the causal problem. It also permits courts to engage in fuzzy-headed thinking about what sort of causal requirement *should* be imposed on plaintiffs, especially in cases that present complications in the availability of causal evidence.³⁵

Richard Wright and Ingeborg Puppe have stated that phrases such as “substantial factor” “are of no help in resolving the causation issue but rather

32. Richard W. Wright, *The Grounds and Extent of Legal Responsibility*, 40 SAN DIEGO L. REV. 1425, 1441 (2003) [hereinafter Wright, *Grounds and Extent*]; Richard W. Wright, *Once More into the Bramble Bush: Duty, Causal Contribution, and the Extent of Legal Responsibility*, 54 VAND. L. REV. 1071, 1132 (2001) [hereinafter Wright, *Once More into the Bramble Bush*]; see Wright, *supra* note 6, at 1788–1803.

The NESS test and the Restatement Third approach are based on the idea of causal sets. The NESS test, developed by Hart and Honoré and “refined and popularized” by Richard Wright, stems from scholarship in philosophy. RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL & EMOTIONAL HARM § 26 cmt. c (AM. LAW INST. 2010). In Hart and Honoré’s *Causation in the Law*, the authors discuss the idea of a “causally relevant factor,” which is a condition “necessary just in the sense that it is one of a set of conditions jointly sufficient for the production of the consequence: it is necessary because it is required to complete this set.” H.L.A. HART & TONY HONORÉ, *CAUSATION IN THE LAW* 112–13 (2d ed. 1985). This concept “relied on [John Stuart] Mill’s idea that a fully described causal law lists all the conditions that together are necessary and sufficient for the occurrence of a certain consequence.” G’sell, *supra* note 24, at 516–17. In 1965, John Mackie “employed an acronym, INUS (for ‘insufficient but non-redundant [necessary] part of an unnecessary but sufficient condition’), to refer to the conditions that make up the minimally sufficient set of abstract conditions that constitute a causal law.” Wright & Puppe, *supra* note 10, at 469. “If applied as a criterion for being an actual causal condition in a concrete singular instance, the INUS criterion would be one way of describing the least stringent, *weak* sense of necessity, which merely requires that a condition be necessary for the sufficiency of a set of actual conditions that was sufficient for the occurrence of the effect, rather than being always necessary (*strict* necessity) or necessary for the effect in the singular instance (*strong* necessity).” *Id.* at 469–70. Wright ultimately sharpened Hart and Honoré’s account in the NESS test, which “confirms causal contribution by each fire” in the two-fire problem. Wright, *supra* note 6, at 1790–91. “Each fire was necessary for the sufficiency of a set of actual antecedent conditions that did not include the other fire.” *Id.* See generally SANDY STEEL, *PROOF OF CAUSATION IN TORT LAW* 15–47 (2015).

33. See Fischer, *supra* note 31, at 277.

34. DOBBS ET AL., *supra* note 1, at 322 (citing cases).

35. Joseph Sanders, Michael D. Green & William C. Powers, *The Insubstantiality of the “Substantial Factor” Test for Causation*, 73 MO. L. REV. 399, 430 (2008) (the authors note that “[b]ecause of other obligations, [William] Powers is only able to join as a co-author of Part I of [the] Article” (*id.* at 399)).

are merely labels applied to an unexplained conclusion.”³⁶ According to the authors, “[t]he words ‘factor,’ ‘contribution,’ and ‘causation’ merely restate the causal issue. The ‘substantial,’ ‘material’ and ‘common sense’ qualifiers only make things worse by adding tests of significance that confuse the causation issue with the normative responsibility issue.”³⁷ In summary, the substantial-factor test is ill-defined, and it arguably leaves the causation inquiry to the intuition of the factfinder.

Nevertheless, the substantial factor test is often viewed as “justified” as applied to cases involving multiple sufficient causes.³⁸ Meanwhile, the courts’ use of alternative standards has not slowed attempts to explain or “modify” the but-for standard “to rescue [it] from its inadequacies.”³⁹ One common “modification” involves

add[ing] details to the description of the injury, including not only the time and/or location at which it occurred but also tautological references to the causal process by which it occurred. For example, in the two-fires situation, the injury is described as the destruction of the house (i) at this particular time, (ii) with this particular debris pattern, and/or (iii) by two fires.⁴⁰

These explanations have, however, been viewed as unsatisfying.⁴¹

A second category of explanations involves affording causal status based on an aggregation of multiple events.⁴² But these explanations have also proved unsatisfying. According to Wright and Puppe:

36. Wright & Puppe, *supra* note 10, at 480–81.

37. *Id.* In an innovative article, J. Shahar Dillbary argued that, under certain circumstances, “group causation theories,” including the “concerted action, substantial factor, and alternative liability” theories, “reduce the parties’ incentives to take care and may result in more, not fewer, injuries.” J. Shahar Dillbary, *Causation Actually*, 51 GA. L. REV. 1, 3–5 (2016). Dillbary used this reasoning to “challenge[] the consensus that group causation theories abandon the actual causation requirement.” *Id.* at 6. In particular, he relied on a “multi-party dynamic” where, under certain circumstances, a party would not have incentive to engage in tortious behavior alone but would have such incentive if joined with the tortious behavior of another, given group causation theories that render the behavior liable. *See id.* at 5–7.

38. DOBBS ET AL., *supra* note 1, at 322–23; *see* Malone, *supra* note 18, at 88–91; Robertson, *supra* note 21, at 1776–77.

39. Wright, *supra* note 6, at 1781.

40. Wright & Puppe, *supra* note 10, at 475.

41. In addition to philosophical criticisms, they have been rejected on the ground that “[i]n law, the required causal relation is between the wrongful aspect of the defendant’s conduct and the properly described legal injury, which usually does not include its specific timing or location. Instead those details serve merely to identify the specific event or state of affairs for which causation of the legally relevant properties (the required legal injury) is at issue.” *Id.* at 476.

42. *Id.* at 477; *see* KEETON ET AL., *supra* note 1, § 41, at 268; *see also* Wright, *supra* note 6, at 1780–81.

Some . . . treat the aggregate condition as a cause while illogically denying causation by any of the individual included conditions. Others illogically treat causation by the aggregate condition as establishing causation by each included condition. In Germany this is known as the “formula of alternatives” (*Alternativenformel*): “Of several conditions which can be eliminated separately but not cumulatively without the effect failing to occur, each is a cause.” Although advertised as an extension of the strong necessity criterion, this formulation instead directly contradicts it. The formula does not merely allow that the alternative cause is not a necessary condition, it prohibits it from being a necessary condition, to avoid being able to declare every fact as a cause for any event by combining it with a fact that is really a cause of it.⁴³

Furthermore, scholars have objected to combining a general application of the but-for standard with alternative standards to be applied in multiple-sufficient-cause situations. Such “combination” approaches allow “policy considerations” to permeate what is intended to be a factual inquiry. In 1956, Wex Malone wrote: “For nearly a century judges and writers have struggled to unravel the tangled skein of fact and policy. Even today the search is on for a judging and language technique that will enable courts to deal with these two components separately and effectively.”⁴⁴ Accordingly, “legal science began to recognize two separate notions—cause-in-fact, and ‘proximate’ or ‘legal’ cause.”⁴⁵ This division is perhaps highlighted most sharply today in

43. Wright & Puppe, *supra* note 10, at 477–78 (internal citations omitted). *See generally* STEEL, *supra* note 32, at 15–47 (“German law also clearly accepts causation in independent sufficiency cases. . . . The BGH recently described [duplicative-causation cases] as follows: ‘If an injury has been caused by several, simultaneously or concurrently effective, events, each of which would have been sufficient on its own to produce the entire injury, then, according to the case law of this court, each event is to be classed as a cause of the injury, as a matter of law, even though none satisfies the *conditio sine qua non* test.’” *Id.* at 23 (quoting BGH NJW 2013, 2018).).

44. Malone, *supra* note 18, at 60.

45. *Id.* *See generally* HART AND HONORÉ, *supra* note 32, at 90 (“[W]hether or not the harm would have happened without the act . . . ‘cause in fact’ . . . is the sole point of contact with what causation means apart from the law. All the remaining components are questions of the law’s policy, to be found in the court’s conception of what limitations are just and expedient or in accord with the rationale or ‘purpose’ of legal rules.”). Note that Hart and Honoré’s concept is intended to be grounded in (non-legal) “common-sense” causal principles rather than in scientific or philosophic notions of cause and effect in particular. *See* H.L.A. HART & TONY HONORÉ, CAUSATION IN THE LAW 1 (1st ed. 1959) [hereinafter HART & HONORÉ 1st ed.] (“[T]he assertion often made by the courts, especially in England, that it is the plain man’s notions of causation (and not the philosopher’s or the scientist’s) with which the law is concerned, seems to us to be true.”). Their account, however, has been criticized as introducing “criteria [that] are neither policy-neutral nor causal.” Wright, *supra* note 6, at 1746; *see also* Richard A. Epstein, *A Theory of Strict Liability*, 2 J. LEGAL STUD. 151, 162 (1973) (“[Hart and Honoré’s] definition, and its

the Restatement Third's rejection of the term "proximate cause" in favor of the term "scope of liability."⁴⁶ But it has been forcefully argued that cause-in-fact, and not only proximate cause, entails policy choices; and exceptions to the but-for test are primary (although arguably not exclusive) reasons.⁴⁷

Although different theories of tort law may place varying degrees of emphasis on the importance of cause-in-fact,⁴⁸ it is commonly believed that "policy considerations have no role to play in the determination of cause-in-fact, 'because no policy can be strong enough to warrant the imposition of liability for loss to which the defendant's conduct has not *in fact* contributed.'"⁴⁹ Furthermore, "[t]he [but-for] test [in particular] reflects a deeply rooted belief that a condition cannot be a cause of some event unless it is, in some sense, necessary for the occurrence of the event. This view is shared by lawyers, philosophers, scientists, and the general public."⁵⁰ But, then, substituting the substantial-factor test, or another altogether different standard, for the counterfactual model, which is built on this idea of the

careful explication . . . have been rejected for the most part in the legal literature . . ."). *See generally* Jane Stapleton, *Choosing What We Mean by "Causation" in the Law*, 73 MO. L. REV. 433, 458–59 (2008).

46. RESTATEMENT (THIRD) OF TORTS: PHYSICAL & EMOTIONAL HARM § 29 cmts. b, g (AM. LAW INST. 2010).

47. *See* Malone, *supra* note 18, at 61–62, 88 ("even with reference to this issue of simple cause the mysterious relationship between policy and fact is likely to be in the foreground"; "[t]he very fact that a new definition of cause is needed in many situations indicates clearly that the but-for rule does not always meet the policy requirements of law"); James E. Viator, *When Cause-In-Fact is More than a Fact: The Malone-Green Debate on the Role of Policy in Determining Factual Causation in Tort Law*, 44 LA. L. REV. 1519, 1526 (1984) ("Torts commentators and courts generally recognize that the but-for test breaks down [in multiple-sufficient-cause situations]. In these two areas ["omissions and multiple sufficient causes"], the determination of cause-in-fact clearly involves normative policy choices . . .").

48. *See* WILLIAM M. LANDES & RICHARD A. POSNER, *THE ECONOMIC STRUCTURE OF TORT LAW* 229 (Harvard Univ. Press 1987) (asserting that "the idea of causation can largely be dispensed with in an economic analysis of torts"; and arguing that a "scholar does not need these terms; he can approach a case in which causation is an issue by asking how the case should be decided consistently with the Hand formula," that is, with the underlying policy considerations). *But see* Guido Calabresi, *Concerning Cause and the Law of Torts: An Essay for Harry Kalven, Jr.*, 43 U. CHI. L. REV. 69, 84 (1975) ("Generally a causal link between an activity and an injury would be required. It would clearly be unproductive to try to induce a modification in conduct for the purpose of reducing injury costs unless we believed the conduct to be causally linked to those injury costs. To put it another way, how can a person be the cheapest cost avoider of an injury if his actions do not increase the chances that the injury will occur?"). *See generally* William M. Landes & Richard A. Posner, *Causation in Tort Law: An Economic Approach*, 12 J. LEGAL STUD. 109 (1983); Steven Shavell, *An Analysis of Causation and the Scope of Liability in the Law of Torts*, 9 J. LEGAL STUD. 463 (1980).

49. Viator, *supra* note 47, at 1526–27 (quoting J. FLEMING, *THE LAW OF TORTS* 170 (6th ed. 1983)) (remarking that "[a]lthough this maxim commands the general allegiance of even Wex Malone, he realized that its application can be troublesome").

50. Wright, *supra* note 6, at 1775.

necessary condition, in cases in which the counterfactual model is thought to fail, is problematic. As David Fischer remarked, in multiple-sufficient-cause cases, courts encounter an issue of policy: “They can retain the necessary cause requirement and exonerate the wrongdoer. Alternatively, they can dispense with the necessary cause requirement, and impose liability on the basis that defendant’s conduct simply was sufficient (in conjunction with the surrounding circumstances) to produce the result.”⁵¹ In other words, replacing the counterfactual model with a more “flexible” standard when the former is thought to fail abandons a factual inquiry based on a widely-accepted notion of cause and effect in favor of a policy decision aimed simply at obtaining the sought-after outcome. As the Supreme Court has stated, “tort law teaches that alternative and less demanding causal standards”—referring to such standards as “a kind of legal fiction or construct”—“are necessary in certain circumstances to vindicate the law’s purposes.”⁵²

In summary, the purported failure of the counterfactual model of causation when applied to multiple-sufficient-cause situations has caused courts and scholars to abandon the counterfactual model, at least as applied to these situations. As I explain in Part V, this effect has had far-reaching implications, well beyond the two-fire problem and other such situations in the torts context.

III. POTENTIAL OUTCOMES: A FRAMEWORK FOR CAUSE AND EFFECT

In this Part, I summarize the basic concepts underlying the potential outcomes framework and then discuss implications for the but-for standard and factual causation generally.

A. *Defining and Estimating Causal Effects*

Causal inference begins with formulating a precise definition of cause and effect. Consider circumstances in which an individual consults a doctor for

51. David A. Fischer, *Causation in Fact in Omission Cases*, 1992 UTAH L. REV. 1335, 1345 (1992) (explaining that “[t]here are two reasons for using these alternative tests instead of the but-for test. First, the but-for test, as applied to this situation, seems to reach the wrong result. People reject the but-for test’s result (i.e., that neither of the twin fires caused the harm) because they intuitively believe that both fires in fact contributed to the destruction of the [lodge]. Second, the but-for test frustrates corrective justice considerations. Imposing liability avoids the unfairness of using the test’s cold logic to exonerate an identified wrongdoer at the expense of the innocent victim.” *Id.* at 1346.).

52. *Paroline v. United States*, 572 U.S. 434, 452 (2014). See generally Wright, *supra* note 6, 1738–39; Malone, *supra* note 18; HART & HONORÉ, *supra* note 32; HART & HONORÉ 1st ed., *supra* note 45; Epstein, *supra* note 45, at 162–63.

backpain. Let us assume that the doctor either provides pain medication or does not, and that the patient's backpain will either persist or cease to persist after an hour.⁵³ If we are interested in making conclusions regarding the causal effect of the pain medication on the patient's backpain, we begin by defining a few fundamentals. First, we can define the patient as an experimental "unit," which is something or someone at a particular point in time that we intend to expose to a "treatment."⁵⁴ A treatment is "an action or intervention that can be initiated or withheld from that unit" at that particular point in time.⁵⁵ In this scenario, let us define two treatments: the administration of pain medication (the "active" treatment) and the administration of no pain medication (the "control" treatment).⁵⁶ Next, we can define the patient's backpain as an "outcome variable" of interest; and we can then define two "potential outcomes," one associated with the administration of pain medication, denoted by $Y(\text{medication})$, and one associated with the administration of no medication, denoted by $Y(\text{no medication})$. A potential outcome is the outcome that would be realized if a unit receives a specific treatment—here, medication or no medication.⁵⁷ Therefore, there are four possibilities: $Y(\text{medication}) = \text{backpain}$, $Y(\text{medication}) = \text{no backpain}$, $Y(\text{no medication}) = \text{backpain}$, $Y(\text{no medication}) = \text{no backpain}$.⁵⁸ We can then define the "unit-level" causal effect as a comparison between these potential outcomes.⁵⁹

Specifically, consider four possible comparisons based on the above potential outcomes: 1) $Y(\text{medication}) = \text{backpain}$ vs. $Y(\text{no medication}) = \text{backpain}$; 2) $Y(\text{medication}) = \text{no backpain}$ vs. $Y(\text{no medication}) = \text{backpain}$;

53. See generally GUIDO W. IMBENS & DONALD B. RUBIN, CAUSAL INFERENCE FOR STATISTICS, SOCIAL, AND BIOMEDICAL SCIENCES: AN INTRODUCTION 3–30 (2015).

54. See *id.* at 5–6.

55. Donald B. Rubin, *The Design Versus the Analysis of Observational Studies for Causal Effects: Parallels with the Design of Randomized Trials*, 26 STAT. MEDICINE 20, 22 (2007); see IMBENS & RUBIN, *supra* note 53, at 4.

56. The "more active" treatment is sometimes referred to as the "active treatment" whereas the "more passive" treatment is sometimes referred to as the "control treatment." These "treatment levels" are often referred to simply as "treatment" and "control," respectively, if doing so would not cause confusion. IMBENS & RUBIN, *supra* note 53, at 4.

57. *Id.* at 3–5. See generally D. James Greiner, *Causal Inference in Civil Rights Litigation*, 122 HARV. L. REV. 533 (2008).

58. Throughout this article, I use terminology and notation based on the Rubin Causal Model. See Donald B. Rubin, *Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies*, 66(5) J. ED. PSYCH. 688 (1974) [hereinafter Rubin, *Estimating Causal Effects*]; Donald B. Rubin, *Bayesian Inference for Causal Effects: The Role of Randomization*, 6 ANNALS STAT. 34 (1978) [hereinafter Rubin, *Bayesian Inference*]; Tirthankar Dasgupta et al., *Causal Inference from 2K Factorial Designs by Using Potential Outcomes*, 77(4) J. ROYAL STAT. SOC'Y SERIES B (STAT. METHODOLOGY) 727 (2015); IMBENS & RUBIN, *supra* note 53.

59. See IMBENS & RUBIN, *supra* note 53, at 5–7.

3) $Y(\text{medication}) = \text{backpain}$ vs. $Y(\text{no medication}) = \text{no backpain}$; 4) $Y(\text{medication}) = \text{no backpain}$ vs. $Y(\text{no medication}) = \text{no backpain}$.⁶⁰ The first and fourth comparisons indicate no causal effect of medication whereas the second and third comparisons indicate a causal effect of medication—the second indicating pain relief and the third indicating prevention of pain relief.⁶¹

In many situations, it may be more appropriate to measure the backpain, for example, using a scale of one to ten, rather than a binary measure. In this case, each of the two potential outcomes— $Y(\text{medication})$ and $Y(\text{no medication})$ —would equal one of multiple values, and the causal effect would again be defined in terms of a comparison between the two potential outcomes. For example, as summarized in Table 1, assuming a constant additive treatment effect and the potential outcomes $Y(\text{medication}) = 4$ and $Y(\text{no medication}) = 8$, the causal effect of medication would be -4 units of pain.⁶²

Table 1. Unit-Level Causal Effect

Unit	$Y(\text{medication})$	$Y(\text{no medication})$	Causal Effect $Y(\text{medication}) - Y(\text{no medication})$
1	4	8	-4

Remember, there is no question here whether this difference, -4 , is due to randomness or a true effect. Rather, we are assuming knowledge of the potential outcome under each level of treatment, medication or no medication. Similarly, it is important to understand that we are *not* conducting a comparison of values before and after medication or a comparison between the pain in one individual who has been given medication and one who has not. Rather, “the causal effect is the comparison of potential outcomes, for the same unit, at the same moment in time post-treatment.”⁶³

Indeed, these cautions highlight “[t]he fundamental problem facing inference for causal effects”: “at most [only] one of the potential outcomes can be realized and thus observed.”⁶⁴ In the example above, the doctor either administers medication or she does not; we therefore observe the potential outcome associated with one treatment level or the potential outcome associated with the other treatment level, but not both. In actuality, in Table

60. *Id.*

61. *Id.*

62. *See generally id.*

63. *Id.* at 6.

64. *Id.* at 6; Rubin, *Bayesian Inference*, *supra* note 58, at 38.

1, the data point in either the second or third column must not exist, and therefore, the unit-level causal effect in the fourth column cannot be observed. For this reason, under the Rubin Causal Model, causal inference is said to be a “missing data problem”: “given any treatment assigned to an individual unit, the potential outcome associated with any alternate treatment is missing.”⁶⁵

Describing causal inference as a missing data problem highlights the important point that causal effects cannot be observed, but rather must be inferred. They must be *estimated*. In the example above, we can observe one of two potential outcomes—the one associated with the level of treatment that is in fact applied to the unit. This is insufficient to determine a causal effect; but causal effects may be inferred by inferring unobserved potential outcomes.⁶⁶

In general, to estimate a causal effect, we can use “replication.” That is, we can test the intervention (e.g., medication) in an “experiment” by exposing multiple units to different treatments.⁶⁷ For each replication, we can observe the potential outcome associated with the assigned treatment, but not those associated with alternative assignments. But now, armed with multiple data points, we can impute *missing* data points and estimate causal effects.

Consider, for example, an experiment to test the effect of the pain medication in the example above. Table 2 illustrates possible results for such an experiment. As described, for each unit, we can observe one potential outcome. For unit 1, who was assigned the treatment “medication,” we observe a pain level of 4. Because she was assigned “medication” rather than “no medication,” we cannot observe her pain level had she received no medication; and therefore, we cannot observe the unit-level causal effect for unit 1. For unit 2, who was assigned the treatment “no medication,” we observe a pain level of 7. We cannot, however, observe his pain level under the counterfactual in which he had received “medication.” And so on and so forth.

Table 2. Causal Inference: A Missing Data Problem

Unit	Y(1=medication)	Y(0=no medication)	Causal Effect
1	4	?	?
2	?	7	?
3	?	9	?

65. IMBENS & RUBIN, *supra* note 53, at 14; *see also* Rubin, *Estimating Causal Effects*, *supra* note 58.

66. *See* IMBENS & RUBIN, *supra* note 53, at 14.

67. *See id.* at 3–30.

4	3	?	?
...
n	3	?	?
	$\bar{Y}^{obs}(1) = 3.3$	$\bar{Y}^{obs}(0) = 8$	$\bar{Y}^{obs}(1) - \bar{Y}^{obs}(0) = -4.7$

To estimate the causal effect, we seek to impute the missing data. A convenient (but not always suitable) approach is to use the mean of the observations for each treatment to impute the missing data for that treatment. For example, in Table 2, we can use $\bar{Y}^{obs}(1) = 3.3$ to impute the missing values for $Y(\text{medication})$ and $\bar{Y}^{obs}(0) = 8$ to impute the missing values for $Y(\text{no medication})$, where $\bar{Y}^{obs}(1)$ refers to the mean of the observed potential outcomes across the units when assigned to “medication” and $\bar{Y}^{obs}(0)$ refers to the mean of the observed potential outcomes across the units when assigned to “no medication.” Once we have imputed the missing potential outcomes, we can estimate the causal effect based on a specified definition. For example, we might have defined the “population-level” causal effect to be the additive difference between the mean of the potential outcomes associated with the treatment level “medication” ($\bar{Y}(1)$) and the mean of the potential outcomes associated with the treatment level “no medication” ($\bar{Y}(0)$), which, in Table 2, we could *estimate* to be -4.7 using the means of the *observed* potential outcomes.⁶⁸ There are well-established methods of inferring causal effects both at the population level and at the unit level.⁶⁹ Note also that implicit in this and other causal inferences under the Rubin Causal Model are two important assumptions: 1) that the treatment assigned to one unit does not affect another unit’s outcome; and 2) that there are no different, or “hidden,” versions of each treatment level.⁷⁰

Thus, as illustrated in the example above, using these assumptions we can estimate causal effects through replication. But, as mentioned earlier, any replication will necessarily be imperfect, since a unit exists only at a single point in time. For example, if we assign an individual to “medication” on Monday and assign the same individual to “no medication” on Tuesday, and compare the pain levels for each, this comparison involves two units: the individual on Monday and the individual on Tuesday. Similarly, if we use identical twins and assign one to “medication” and the other to “no

68. For a description of Jerzy Neyman’s “repeated sampling” approach, see IMBENS & RUBIN, *supra* note 53, at 83–84.

69. For a recent experiment using “Fisher’s method” and the potential outcomes framework, see Hillel J. Bavli & Reagan Mozer, *The Effects of Comparable-Case Guidance on Awards for Pain and Suffering and Punitive Damages: Evidence from a Randomized Controlled Trial*, 37 YALE L. & POL’Y REV. 405 (2019). The Rubin Causal Model allows inference using Neyman’s approach or Fisher’s approach. See Rubin, *Bayesian Inference*, *supra* note 58.

70. IMBENS & RUBIN, *supra* note 53, at 9–13.

medication” at the same point in time, this comparison also involves two units: each twin at the same point in time. The replication is necessarily imperfect, since even identical twins differ to some degree, as does an individual on Monday compared to the same individual on Tuesday—let alone, different (non-twin) individuals at different times.

Imperfect replication can degrade the imputation of missing potential outcomes and the estimation of causal effects. Specifically, if the background characteristics of the different units, such as age, gender, race, etc.—variables known as “covariates”—differ from one unit to the next, we do not know whether the difference in outcome (for example, pain level) is due to the difference in treatment or the difference in covariates.⁷¹ For example, if subjects in the above experiment are recruited from a population of young adults, and these subjects are permitted to choose between receiving medication for backpain or receiving no medication, then a comparison between the average pain level of those who received medication and the average pain level of those who received no medication may be misleading. For example, if the medication contains a substance known to be harmful to fetuses or breastfed infants, the units in the “no medication” group may be disproportionately women, and the units in the “medication” group disproportionately men, because a proportion of the population from which the subjects were recruited consisted of women who were pregnant or breastfeeding and did not want to risk injuring their fetuses or infants with the medication. Meanwhile women (and pregnant or breastfeeding women in particular) may have different pain thresholds than men, or perhaps may rate pain differently than men; therefore, any observed effect may be due not to the medication, but to differences in covariates, and specifically proportions of men and women, in each treatment group.

Furthermore, regardless whether the units in the “no medication” group are disproportionately women, they may consist of subjects with disproportionately low pre-treatment pain levels (also a covariate like race and gender). After all, subjects with high pain levels may be more likely to choose medication over no medication. As above, any observed effect may therefore be due not to the medication, but to differences in pre-treatment pain levels in the two treatment groups.

This example illustrates the importance of “covariate balance” across treatment groups. It also illustrates the importance of a concept known as the

71. A covariate is a background characteristic that cannot be affected by the treatment condition to which a unit is assigned. For example, whether a unit is assigned to “medication” or “no medication” will not change the unit’s gender. On the other hand, the unit’s temperature could be affected by whether the unit receives medication or no medication and is therefore not a covariate. *See id.* at 15–16.

“assignment mechanism,” which determines how each unit “[comes] to receive the treatment level actually received.”⁷²

To achieve covariate balance in the experimental context, a researcher may seek to use a “randomized experiment,” in which the treatment assignments are random. Often, however (including in many torts contexts), randomization is not possible. For example, for a case involving the question whether a botched heart surgery of a certain type caused heart failure, a researcher could not ethically conduct a randomized experiment in which she would randomize which patients receive botched heart surgeries and which do not in order to determine the effect of botched surgeries on heart failure. Similarly, if it is known that exposure to asbestos may have a range of harmful effects, a researcher likely could not randomize exposure to asbestos to test its effects on human health. Further, even under circumstances in which a randomized experiment is ethically permissible, it may be ruled out due to the high costs that such studies entail.

Randomized experiments are the “gold standard” for making causal inferences.⁷³ But in circumstances in which randomization is not possible, it may be appropriate to conduct an “observational study,” which is a study in which the researcher does not control the assignment of treatments to units.⁷⁴ For example, the researcher may use existing data regarding exposures to asbestos and associated health outcomes. The problem with observational studies is that, because the researcher is unable to control the assignment mechanism, it is difficult if not impossible to determine whether there is covariate balance, and therefore whether an observed disparity in an outcome variable is due to a causal effect or to a covariate imbalance. Indeed, a well-designed observational study will often attempt to “approximate, or attempt to replicate, a randomized experiment.”⁷⁵ This may involve comparing the outcomes of units that are similar to each other—that is, that have similar covariates—but that receive different treatment conditions. This is a challenging exercise, and, even if possible, it will be difficult to ascertain that all relevant variables have been identified and balanced. One method of

72. *Id.* at 13–15.

73. Rubin, *supra* note 55, at 20.

74. See IMBENS & RUBIN, *supra* note 53, at 31, 41–42. An observational study can be described as “an empiric investigation in which ‘the objective is to elucidate cause-and-effect relationships . . . [in which] it is not feasible to use controlled experimentation, in the sense of being able to impose the procedures or treatments whose effects it is desired to discover, or to assign subjects at random to different procedures.’” PAUL R. ROSENBAUM, OBSERVATION AND EXPERIMENTATION: AN INTRODUCTION TO CAUSAL INFERENCE 65 (2017) (quoting William G. Cochran, *The Planning of Observational Studies of Human Populations (with Discussion)*, 128 J. ROYAL STAT. SOC’Y, SERIES A (GEN.) 234, 234 (1965)).

75. Rubin, *supra* note 55, at 25.

simultaneously balancing numerous covariates is to use a device called “propensity scores,” which provide the probability that a unit will receive a certain treatment, given the unit’s covariates.⁷⁶ A propensity score can be understood as a summary value reflecting a unit’s array of covariate values.⁷⁷

Covariate balance, however, is not the only reason to design an observational study to replicate a randomized experiment. A major theme underlying the potential outcomes framework is the importance of carefully defining causal questions and causal effects. This entails precision in defining “primitive concepts,” such as units, treatments, and outcome variables,⁷⁸ as well as estimands—the objects we want to estimate (e.g., a causal effect)—and estimators—the objects that we intend to use to estimate the estimands in light of imperfect information (e.g., a formula based on average observed outcomes). A primary purpose for designing an observational study to replicate a randomized experiment is thus to facilitate good study design with well-defined causal effects.

Furthermore, it is important to understand that the potential outcomes framework is more than simply a set of methods for analyzing data. Indeed, although the discussion of “replication” above relates to methods for inferring causal effects from data—methods that are useful in many torts contexts involving questions of cause and effect—a central feature of the potential outcomes framework is its simple and intuitive structure for *defining* cause and effect. It is a framework for conceptualizing causal questions and causal effects, regardless whether the method to be applied for inferring causal effects is based on replication and a formal dataset or simply on a trier of fact that draws inferences from the evidence presented at a trial.

B. *Implications for But-For Causation in Multiple-Sufficient-Cause Situations*

In this section, I explain the relationship between but-for causation and the potential outcomes framework and implications of this relationship for the role of but-for causation for resolving causal questions involving multiple sufficient causes.

Factual Causation and Scientific Cause and Effect. Factual causation is intended to capture the meaning of “actual,” or “scientific,” cause and

76. See Paul R. Rosenbaum & Donald B. Rubin, *The Central Role of the Propensity Score in Observational Studies for Causal Effects*, 70 *BIOMETRIKA* 41, 41 (1983).

77. See generally *id.*; Rubin, *supra* note 55. See also Greiner, *supra* note 57, at 574–75.

78. Greiner, *supra* note 57, at 558.

effect.⁷⁹ It is conceptualized as such, and, in many cases (such as toxic tort cases), courts accept and sometimes *require* (or at least hold as the “gold standard”) proof of causation through statistical evidence establishing a causal connection between a defendant’s misconduct and a plaintiff’s injury.⁸⁰ Indeed, the factual causation inquiry can be understood as an

79. See DOBBS ET AL., *supra* note 1, at 313–14 (referring to “factual cause problem[s] [that] center[] on scientific doubt, or at least on lay ignorance about the connection between the defendant’s acts and the plaintiff’s injury”; “[t]hese causal issues raise questions of fact in the scientific sense”; “the plaintiff must . . . present evidence that causation is not merely scientifically possible, but that it existed in her particular case”); KEETON ET AL., *supra* note 1, § 41, at 264–65 (“This question of ‘fact’ ordinarily is one upon which all the learning, literature and lore of the law are largely lost. It is a matter upon which lay opinion is quite as competent as that of the most experienced court.”); see also *Burrage v. United States*, 571 U.S. 204, 210–11 (2014) (“A thing ‘results’ when it ‘[a]rise[s] as an effect, issue, or outcome *from* some action, process or design.’ ‘Results from’ imposes, in other words, a requirement of actual causality. ‘In the usual course,’ this requires proof “that the harm would not have occurred” in the absence of—that is, but for—the defendant’s conduct.” (internal citations omitted)); Lawrence Crocker, *A Retributive Theory of Criminal Causation*, 1994 J. CONTEMP. LEGAL ISSUES 65, 67 (1994) (“Judges comment from time to time on how difficult is the concept of legal or proximate causation in comparison to the straightforward concept variously referred to as ‘cause in fact’ or ‘scientific’ or ‘but for’ causation. Philosophers of science, by contrast, are inclined towards the view that scientific causation or cause in fact is terribly difficult.”) (citing sources); Viator, *supra* note 47, at 1523 (“The Green methodology first confines the causation issue to a neutral, purely scientific inquiry, namely, cause-in-fact or what Green calls either causal relation or causal connection.”).

80. See *Norris v. Baxter Healthcare Corp.*, 397 F.3d 878, 882 (10th Cir. 2005) (“epidemiology is the best evidence of general causation in a toxic tort case”; “where epidemiology is available, it cannot be ignored”); *Brock v. Merrell Dow Pharm., Inc.*, 874 F.2d 307, 311 (5th Cir. 1989) (“Undoubtedly, the most useful and conclusive type of evidence in a case such as this is epidemiological studies.”); *Pozefsky v. Baxter Healthcare Corp.*, No. 92CV0314LEKRW, 2001 WL 967608, at *3 (N.D.N.Y. Aug. 16, 2001) (“In the mass torts context, epidemiology is the best evidence of causation.”); *In re Breast Implant Litig.*, 11 F. Supp. 2d 1217, 1224 (D. Colo. 1998) (“[e]pidemiology is the best evidence of causation in the mass torts context”; “epidemiological studies are necessary to determine the cause and effect between breast implants and allegedly associated diseases”); see also *In re Agent Orange Prod. Liab. Litig.* MDL No. 381, 818 F.2d 187 (2d Cir. 1987); *Burton v. Wyeth-Ayerst Labs. Div. of Am. Home Prod. Corp.*, 513 F. Supp. 2d 719, 730 (N.D. Tex. 2007) (citing *Merrell Dow Pharm., Inc. v. Havner*, 953 S.W.2d 706, 715 (Tex. 1997)) (stating that when “direct” evidence of specific causation “is unavailable, claimants may attempt to demonstrate an increased risk posed by exposure to the substance in question in an effort to establish causation; this is generally done through epidemiological studies”); *In re Agent Orange Prod. Liab. Litig.*, 611 F. Supp. 1223, 1231 (E.D.N.Y. 1985) (discussing epidemiological studies) *aff’d sub nom. In re Agent Orange Prod. Liab. Litig.* MDL No. 381, 818 F.2d 187 (2d Cir. 1987); *Havner*, 953 S.W.2d at 715 (“[W]hen the incidence of a disease or injury is sufficiently elevated due to exposure to a substance, someone who was exposed to that substance and exhibits the disease or injury can raise a fact question on causation.”); David H. Kaye & David A. Freedman, *Reference Guide on Statistics*, in REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 211, 218 (3d ed. 2011) (explaining that “randomized controlled experiments”—the gold standard for answering questions of causality—“are ideally suited for demonstrating causation”). (Note that the Reference Manual on

observational study designed by the court based on the law, in which a trier of fact draws inferences from evidence and makes causal conclusions.⁸¹

To be sure, factual causation is frequently described as causation in the “ordinary,” “plain,” or “common” sense of the term. But it is important to understand that this “ordinary” meaning of causation correlates extremely well with, if it is not altogether equivalent to, the scientific meaning of cause and effect. Like the scientific meaning, the “ordinary” meaning captures the idea of the counterfactual, and of the but-for concept in particular, as has been confirmed repeatedly by the Supreme Court.⁸²

Furthermore, to determine the “ordinary” meaning of causation, we can examine entries for the term in well-established dictionaries. Such entries frequently serve as the Supreme Court’s primary source when interpreting the “ordinary meaning” of statutory language. In particular, to interpret a statute, the Supreme Court “begin[s] with the language employed by Congress and the assumption that the ordinary meaning of that language accurately expresses the legislative purpose”;⁸³ in determining the “ordinary meaning” of the statutory language, the Court examines entries for the terms at issue in well-established dictionaries such as Merriam-Webster and Oxford Dictionaries.⁸⁴

Scientific Evidence uses the term “causation” to refer to “the task of attributing cause and effect” in both legal and scientific contexts. Jerome P. Kassirer & Gladys Kessler, *Preface* to REFERENCE MANUAL ON SCIENTIFIC EVIDENCE xiii, xiii–xiv (3d ed. 2011).) *See generally* David Rosenberg, *The Causal Connection in Mass Exposure Cases: A ‘Public Law’ Vision of the Tort System*, 97 HARV. L. REV. 849 (1984) (discussing “weak” and “strong” rules). The potential outcomes framework itself has been used in litigation and has appeared or has been the subject of discussion in a number of law review articles. *See, e.g.*, Jonah B. Gelbach, *Locking the Doors to Discovery? Assessing the Effects of Twombly and Iqbal on Access to Discovery*, 121 YALE L. J. 2270 (2012); Greiner, *supra* note 57 (applying the potential outcomes framework to the civil rights litigation context); Daniel E. Ho & Donald B. Rubin, *Credible Causal Inference for Empirical Legal Studies*, 7 ANN. REV. L. & SOC. SCI. 17 (2011). *See also* Donald B. Rubin, *Estimating the Causal Effects of Smoking*, 20 STAT. MEDICINE 1395, 1410–12 (2001) (analyzing causal effects of alleged misconduct of the tobacco industry, and discussing apportionment of damages in circumstances involving alleged misconduct of tobacco and asbestos industries).

81. The potential outcomes framework, which integrates concepts such as causal estimands and causal estimators, unit-level causal effects, and population-level causal effects, would arguably facilitate a tighter integration of legal standards of causation and standards of scientific proof of causation, and perhaps a range of other causation-related concepts—for example, the concepts of “probabilistic” causation and “actual” causation. *See generally* Wright & Puppe, *supra* note 10, at 496–97 (“[Guido Calabresi] also included a ‘causal linkage’ concept, which is more accurately described as a ‘probabilistic linkage,’ since it merely refers to *ex ante* increased risk . . .”).

82. *See, e.g.*, *Burrage*, 571 U.S. at 210–11; *Gross v. FBL Fin. Servs., Inc.*, 557 U.S. 167, 176–77 (2009).

83. *Engine Mfrs. Ass’n v. S. Coast Air Quality Mgmt. Dist.*, 541 U.S. 246, 252 (2004) (internal quotation marks and citations omitted).

84. *See, e.g.*, *Burrage*, 571 U.S. at 210–11; *Gross*, 557 U.S. at 176.

Thus, examining the entries for the term “causation” in these dictionaries, the definitions provided in Merriam-Webster are “the act or process of causing” and “the act or agency which produces an effect”; and in Oxford Dictionaries, “[t]he action of causing something” and “[t]he relationship between cause and effect; causality.”⁸⁵ Most illuminating for our purposes are the “example sentences” in each entry. The first two examples provided by Merriam-Webster are, “the role of heredity in the *causation* of cancer,” and “in a complex situation *causation* is likely to be multiple”; and by Oxford Dictionaries, “the postulated role of nitrate in the causation of cancer,” and “a strong association is not a proof of causation.”⁸⁶ It is clear from these example sentences, as well as the definitions provided, that the dictionary meaning of the term causation, and thus the “ordinary meaning” of the term causation (based on the Supreme Court’s frequent analysis for deriving the ordinary meaning of causal terms and terms generally), is intended to capture the scientific concept of cause and effect.

If factual causation entails scientific cause and effect, then the law should look to the sciences to inform its causal framework. The predominant scientific model (or at least *a* predominant scientific model) for asking and answering questions of cause and effect is the potential outcomes framework. This framework provides an ideal setting in which to conceptualize multiple-sufficient-cause problems and factual causation generally.

But-For Causation and the Counterfactual Model. A central theme of this article is that factual causation cannot be defined simply by a narrow test to be applied rigidly to every situation. Rather, it is a theoretical framework for establishing a defined connection between a defendant’s act and a plaintiff’s injury. Different circumstances may call for different applications of this framework in examining such a connection. For most cases, a rigid “test” may suffice; but complex cases, and specifically, those involving multiple sufficient causes, require additional nuance. This nuance is supplied by the broader counterfactual model, the potential outcomes framework.

The but-for conception of causation fundamentally entails a comparison of counterfactuals corresponding to some defined intervention. In particular, however, but-for causation should be understood as a specific measure of interest, or “estimand,” within the potential outcomes framework. In other words, the potential outcomes model supplies a broad framework for defining and answering causal questions through counterfactual reasoning, one

85. *Causation*, LEXICO, <https://www.lexico.com/en/definition/causation> [<https://perma.cc/GPL5-6QJU>] (last visited Sept. 8, 2019); *Causation*, MERRIAM-WEBSTER, www.merriam-webster.com/dictionary/causation [<https://perma.cc/LD2Z-MCPC>] (last visited Sept. 8, 2019).

86. LEXICO, *supra* note 85; MERRIAM-WEBSTER, *supra* note 85.

specific aspect of which is the but-for measure. But, it is the broader counterfactual model, the potential outcomes framework, and not the specific but-for measure, that the law intends to capture in the factual causation inquiry. After all, this inquiry is intended to be one of “actual,” or “scientific,” cause and effect, which is precisely what the potential outcomes framework represents.⁸⁷

The but-for measure in particular, although logically appropriate as a measure of interest in single-factor problems, is not necessarily a sensible estimand in multiple-sufficient-cause problems, at least not if the law seeks a measure that is consistent with both common and scientific reasoning regarding cause and effect. The potential outcomes framework, however, allows for alternative measures that build on the same counterfactual reasoning that is central to the but-for standard but that are refined for multiple-factor problems. In the following Parts, I introduce these alternatives by describing and applying the “factorial” approach to multiple sufficient causes.

Prior to proceeding to the factorial approach, however, it is important to note that there are two ways to understand applications of the potential outcomes framework to multiple-sufficient-cause problems. Under the first understanding, the potential outcomes framework carries important implications for the but-for standard of causation directly. According to this understanding, the but-for standard is itself intended to entail scientific cause and effect and the broader counterfactual model, but it has been defined by courts too narrowly, such that it is not in fact capturing the full extent of the notions that it entails. The potential outcomes framework, therefore, informs and expands it to entail a broader set of estimands—and specifically, estimands that may be more suitable for resolving multiple-sufficient-cause problems.

Under the second understanding, the but-for standard maintains its narrow meaning and excludes the broader counterfactual reasoning of the potential outcomes framework. Using this understanding, however, the but-for standard should be understood only as a specific measure within the potential outcomes framework and not necessarily applicable to multiple-sufficient-cause situations. As such, the law should look to the broader potential outcomes framework, and not necessarily the but-for measure, to ask and answer causal questions involving multiple sufficient causes.

87. Note, the factual nature of the causal question in the potential outcomes framework is highlighted by Donald Rubin’s frequent remark that causal inference is a missing data problem: the structure of the inquiry places the focus on the inference of unobserved, or “missing,” potential outcomes. *See, e.g.,* IMBENS & RUBIN, *supra* note 53, at 14.

Both of these understandings are valid. The first understanding has the advantage of highlighting that, as applied to multiple sufficient causes, the potential outcomes framework is simply a logical extension of the basic counterfactual concept that gives rise to the but-for standard. The second understanding has the advantage of maintaining the current meaning of but-for causation while placing it in a broader context that allows for a more refined approach to complex causal problems.

IV. A “FACTORIAL” APPROACH TO MULTIPLE SUFFICIENT CAUSES

Whether the counterfactual model of causation can account for cases involving multiple sufficient causes and how it treats these cases have important implications. In this Part, I show that the counterfactual model does not fail in multiple-sufficient-cause situations. Rather, if employed with proper nuance, it yields results that are consistent with intuition and court decisions. I apply a “factorial” approach in the potential outcomes framework to show that both factors in the concurrent multiple-sufficient-cause problem and the first factor in the successive multiple-sufficient-cause problem can be understood as actual causes of the destructive outcomes in those problems.

A. *Concurrent Multiple Sufficient Causes in a Multi-Treatment Setting*

The key to properly applying the counterfactual model to these complex problems is to conceptualize the problems in a way that reflects their complexity. We can accomplish this by applying the potential outcomes framework in a multi-treatment setting. Let us take the two-fire problem as an example. To start, let us use a basic principle of the Rubin Causal Model, discussed above, that observational studies should be designed and conceptualized to replicate a randomized experiment.⁸⁸ We begin by defining

88. See *supra* notes 73–81 and accompanying text. As discussed *supra*, in addition to facilitating balance of covariates across treatment conditions, approaching an observational problem in this fashion, both conceptually and methodologically, allows the researcher to construct the problem rigorously, including carefully defining units, treatments, outcome variables, and causal effects. As Donald Rubin has emphasized, “A crucial idea when trying to estimate causal effects from an observational dataset is to conceptualize the observational dataset as having arisen from a complex randomized experiment, where the rules used to assign the treatment conditions have been lost and must be reconstructed.” Donald B. Rubin, *For Objective Causal Inference, Design Trumps Analysis*, 2 ANNALS APPLIED STAT. 808, 815 (2008). Frequently, rigorous design is lost in observational studies—whether in law or in the sciences. A problem is observed in retrospect and the study is mistakenly designed to reflect that perspective. But our conception of an intervention should not change based on whether a researcher conducts

foundational concepts, such as treatments and outcome variables.⁸⁹ Because there are two factors at issue—Fire A and Fire B—we can adopt a multi-treatment, or “factorial,” design in which we define two treatments. This “factorial” approach to the problem, although not necessary to make the argument herein, provides a clear way of understanding and analyzing multiple-sufficient-cause situations.

The classical but-for approach to the two-fire problem effectively compares the potential outcome associated with the occurrence of the fire at issue to the potential outcome associated with a state of the world in which the fire had not occurred, given the occurrence of the other fire. This approach treats each fire abstractly (i.e., individually) in the sense that it views each fire as a single intervention conditioned on the existence of the other fire. But, approaching the problem as a two-factor, rather than a single-factor, problem allows examination of a broader range of effects, and thereby a more comprehensive understanding of the fires’ roles in destroying the lodge than we obtain using a single-factor design.

When there is only a single factor, the counterfactuals under the but-for conception are straightforward: we ask simply, what would have happened had the intervention not occurred? When there are two factors that are both treated as interventions, however, the counterfactuals are more complex: we want to know what would have happened had the first intervention not occurred *and* what would have happened had the second intervention not occurred. For example, in the concurrent two-fire problem, we consider what would have happened had Fire A not occurred and also what would have happened had Fire B not occurred.

Using this model requires that we further specify what is meant by causation. It can mean different things. For example, concerning the effect of

a prospective experiment or an observational study. “As a consequence of our conceptualization of an observational study’s data as having arisen from a hypothetical randomized experiment, the first activity is to think hard about that hypothetical experiment. To start, what exactly were the treatment conditions and what exactly were the outcome (or response) variables?” *Id.* Note that this approach arguably addresses (or is at least relevant to) a common criticism of the but-for standard—that it captures a multitude of events as causes that are in fact irrelevant to the legal conception of causation. The potential outcomes framework, in a sense, takes a prospective conceptual approach to causality; and this is useful for conceptualizing causation in light of the uncountable events that have led to a plaintiff’s injury. The approach requires carefully defining interventions of interest, as distinct from the infinite events that have possibly led to a hypothesized effect.

89. It is interesting and perhaps useful—although beyond the scope of the current article—to consider whether experimental design and the potential outcomes framework could apply in certain contexts to facilitate a more concrete approach to *proximate* causation by characterizing the inquiry in terms of a randomized experiment that examines the effect of a defined intervention on a defined outcome variable through “samples” of hypothetical worlds that feature either treatment or control conditions.

Fire A, it can refer to a comparison between potential outcomes when Fire A occurs and when Fire A does not occur, conditional on the occurrence of Fire B; or it can refer to an *unconditional* comparison between potential outcomes when Fire A occurs and when Fire A does not occur—in effect, aggregating over potential outcomes when Fire B does and does not occur. These distinct estimands, sometimes referred to as “interaction effects” and “main effects,” respectively, may lead to different conclusions, that is, a finding of “no causation” for one but “causation” for the other.

Let us examine these effects in greater detail by applying a factorial design within the potential outcomes framework. This involves examining the effects of the fires—the factors—on the destruction of the lodge—the outcome variable—under various “treatment combinations.”⁹⁰ Each treatment combination involves a set of factors, where each factor is set to a specific “level.”⁹¹ For example, the lodge in the two fire problem may be exposed to a treatment combination that involves the factors Fire A and Fire B, where Fire A is set to the level “on” and Fire B is set to the level “off.” Using this design, we can study the main effect of each factor on the outcome variable—that is, the effect of each factor aggregated over the levels of the other factor—as well as interaction effects, or the effects of each factor conditional on a specific level of the other factor.⁹² Both sets of effects are based on the counterfactual model; and our intuition regarding causation in multi-factor problems likely derives from both main effects and interaction effects.

Consider the concurrent two-fire problem in the context of a “2x2 factorial design,” involving two factors, Fire A and Fire B, each having two levels, “on” and “off.” Let us define the outcome variable as a binary variable *Y* that equals 1 if the lodge is destroyed and 0 if the lodge is not destroyed.

First, we are interested in knowing the potential outcome associated with each combination of the two factors. The 2x2 table in Figure 1 illustrates these potential outcomes, where, by assumption in the concurrent two-fire problem, each fire is independently sufficient to destroy the lodge.

	Fire A = off	Fire A = on
Fire B = off	Not Destroyed	Destroyed

90. See Dasgupta et al., *supra* note 58, at 727; see also Bavli & Mozer, *supra* note 69, at 420.

91. See Dasgupta et al., *supra* note 58, at 727.

92. See generally *id.*

Fire B = on	Destroyed	Destroyed
-------------	-----------	-----------

Figure 1. 2x2 table illustrating potential outcomes associated with treatment combinations involving two factors, Fire A and Fire B, each having two levels, “on” and “off.”

As Figure 1 illustrates, the lodge is destroyed for the combinations (Fire A = on, Fire B = on), (Fire A = off, Fire B = on), (Fire A = on, Fire B = off), but is not destroyed for the combination (Fire A = off, Fire B = off).⁹³

Distinguish estimands from estimators. In this Part, we are interested in defining estimands. That is, assuming we have all the data we could need (which we indeed have in Figure 1), what value are we interested in knowing—specifically, how should we define a causal effect?

Using numerical values to represent “destroyed” (1) and “not destroyed” (0), we can define interaction and main effects. In these terms, the traditional but-for approach to the concurrent two-fire problem would effectively focus solely on the interaction effects of each fire⁹⁴—for example, determining the interaction effect of Fire A when Fire B is “on” by contrasting the potential outcomes when Fire A is “on” and when Fire A is “off,” given that Fire B is “on.” Using the notation above, we can calculate this interaction effect to be $1 - 1 = 0$. As expected, there is no causal effect of Fire A when conditioning on the existence of Fire B.

But the traditional approach altogether ignores the main effects of each fire. These effects account for two sets of counterfactuals: when Fire A is “on” versus “off” and when Fire B is “on” versus “off.” In particular, the main effect of Fire A on the lodge’s destruction is based on a comparison between an aggregation of both potential outcomes when Fire A is “on” (aggregated over the two possibilities (“on” and “off”) for Fire B) and an aggregation of both potential outcomes when Fire A is “off” (again aggregated over the two possibilities for Fire B). As a simplified example, we can define the unit-level main effect of Fire A as the additive difference between the mean potential outcome when Fire A is “on” and the mean potential outcome when Fire A is “off,” and the unit-level main effect of Fire B as the additive difference between the mean potential outcome when Fire B is “on” and the mean potential outcome when Fire B is “off.”⁹⁵ When determining the effect of one fire, this comparison aggregates over the

93. See generally *id.* at 731–32.

94. In actuality, the traditional approach does not account for both fires simultaneously as interventions and seek to understand how these interventions “interact”; rather it considers each fire individually and assumes that all other features of the problem remain constant, including the existence of the other fire.

95. This is only one of many possible definitions we could consider.

various levels of the other fire. Therefore, using these definitions, the unit-level main effect of Fire A can be written as $(1+1)/2 - (1+0)/2 = 1 - 1/2 = 1/2 = 50\%$. Similarly, using these definitions, the unit-level main effect of Fire B can be written as $1/2$, or 50%.

This particular estimand—i.e., aggregating using the mean—is presented for illustrative purposes only and is not meant as a prescription for a main-effects analysis in multiple-sufficient-cause situations. Indeed, there are many ways one could aggregate over potential outcomes and compare those aggregations when each factor is on versus off. In the current Part, I introduce the main-effects analysis as an alternative within the counterfactual model to the traditional but-for measure, and I explain its logic and applicability. In the following Part, I argue in favor of a particular estimand—and specifically, that this method can be applied in practice through the causal-set approach of the NESS test and the Restatement Third.

Note that, assuming use of the above estimand (employing the mean potential outcome), the values that this main-effects analysis generate (50% and 50%) can be interpreted to yield a legal outcome. One possibility is to interpret them in terms of proportions: as indicating that Fire A and Fire B share equally in causing the destruction of the lodge. It is important, however, not to confuse the percent effect defined in the factorial design with the preponderance-of-the-evidence standard frequently applied to proof of causation. The former does not represent a level of proof. Similarly, it does not represent a percent increase in risk.⁹⁶ Distinguish this result from the statement, “starting a fire is causal in that it increases the risk of destroying the lodge.”⁹⁷ A 50% effect could, in certain contexts, be used to indicate an increase in risk, but this is not how it should be interpreted generally. Our result is not probabilistic: we know with certainty what will happen for each combination of factors, and our result is based entirely on these combinations. Each percent value represents the proportion of combinations of the interventions for which the lodge is destroyed when each intervention (i.e., each fire) is “on” versus “off.” This proportion can then serve as the basis for a legal rule to apportion liability.

96. Courts sometimes fail to distinguish between estimands—the things we want to estimate, such as causal effects—and estimators—what we use to estimate causal effects and other estimands. For example, in determining whether a plaintiff has established beyond a preponderance of the evidence that *x* caused *y*, a court may combine the increased risk of an outcome required to satisfy the preponderance standard and the confidence that the trier of fact should have in the evidence proving that increased risk. In any event, the percentage effect defined in the factorial design is distinct from any percentage used to characterize a standard of proof.

97. See, e.g., Calabresi, *supra* note 48, at 71 (“There is a causal link between an act or activity and an injury when we conclude on the basis of the available evidence that the recurrence of that act or activity will increase the chances that the injury will also occur.”).

For example, the analysis above and resulting proportions could form the basis of legal rules involving apportionment based on causation or even joint and several liability.⁹⁸ For the latter, each fire can be characterized as a cause of an indivisible injury, the destruction of the lodge.

Where the tortious acts of two or more wrongdoers join to produce an indivisible injury, that is, an injury which from its nature cannot be apportioned with reasonable certainty to the individual wrongdoers, all of the wrongdoers will be held jointly and severally liable for the entire damages and the injured party may proceed to judgment against any one separately or against all in one suit.⁹⁹

Relying on a main-effects analysis, each fire can be understood as a cause of the destruction of the lodge, and as having joined with the other fire “to produce an indivisible injury.”¹⁰⁰

Ultimately, defining the unit-level main effect of each fire as the additive difference between the mean potential outcome when the fire is “on” and the mean potential outcome when the fire is “off”—and particularly, employing

98. See generally RESTATEMENT (THIRD) OF TORTS: APPORTIONMENT OF LIAB. §§ 17, 26 (AM. LAW INST. 2000); DOBBS ET AL., *supra* note 1, at 878 (“The principle of causal apportionment can apply between a plaintiff and a defendant as well as between defendants, as where the defendant’s asbestos causes lung damage and the plaintiff’s smoking causes a different lung damage, with both contributing to a shortness of breath. If evidence shows a basis for saying that the asbestos caused 90% of the disability, the defendant will be liable only for that portion of the harm. If no evidence shows a basis for causal apportionment, the court may allocate liability in proportion to fault or responsibility instead, unless special considerations of public policy bar fault apportionment.”); DAN B. DOBBS ET AL., THE LAW OF TORTS § 192 (2d ed. 2018) (“When the plaintiff presents evidence that she suffered a single or indivisible injury at the hands of two or more tortfeasors, the burden is on the party who seeks to avoid responsibility for the entire damages to prove the magnitude of divisible damages. The effect of this rule is that when the plaintiff suffers injuries that are similar in nature or consequences, so that they cannot be separated in any practical way and cannot be attributed separately to the separate tortfeasors, each tortfeasor is treated as a cause of the entire indivisible injury. Under this rule, causal apportionment generally does not apply to indivisible injuries. Instead, a court must either seek an apportionment based on fault or responsibility, or hold multiple defendants jointly and severally liable.”) (citing RESTATEMENT (THIRD) OF TORTS: APPORTIONMENT OF LIAB. § 26 cmt. h (AM. LAW INST. 2000)). See also Rubin, *supra* note 80, at 1410–12 (applying causal framework to apportionment of damages in “joint causation” situation involving combined “misconduct of the asbestos and tobacco industries,” and describing “three counterfactual worlds: the world without misconduct by the tobacco industry, . . . the world without misconduct by the asbestos industry, . . . and the world without misconduct by either industry”).

99. *Landers v. E. Tex. Salt Water Disposal Co.*, 248 S.W.2d 731, 734 (Tex. 1952); see also *In re Modafinil Antitrust Litig.*, 837 F.3d 238, 262 n.29 (3d Cir. 2016), *as amended* (Sept. 29, 2016) (“Under the doctrine of joint and several liability, “[i]f the tortious conduct of each of two or more persons is a legal cause of harm that cannot be apportioned, each is subject to liability for the entire harm, irrespective of whether their conduct is concurring or consecutive.”) (quoting RESTATEMENT (SECOND) OF TORTS § 879 (AM. LAW INST. 1979)).

100. *Landers*, 248 S.W.2d at 734.

the mean to aggregate potential outcomes, as in the illustration above—can lead to various difficulties. Some estimands are better than others. But the point is this: Defining causal effects using main effects rather than interaction effects (and a but-for measure in particular) allows an estimand that accounts for *both* sets of counterfactuals and therefore all four potential outcomes in the 2x2 matrix in Figure 1; and this analysis is a logical and well-accepted application of counterfactual reasoning to define causal effects in the potential outcomes framework, and one that leads to sensible conclusions.

B. Why Employ a Main-Effects Analysis: A Logical Explanation

The above analysis leads to results that are consistent with intuition and court decisions without resorting to exceptions to the counterfactual model.¹⁰¹ But an obvious question is this: the factorial approach models causal effects as potentially based on main effects, interaction effects, or perhaps both; but if it is known that two fires in fact occurred, why, when determining the effects of either of the two fires, should we not automatically condition on the occurrence of the other fire? Even in the context of a multi-factor design, why should we not examine only *interaction* effects, thus conditioning on the existence of the other fire, which we know to have been “on”?

To explain the intuition behind this approach, and, in particular, behind examining the main effect of each fire, I would ask the reader to consider the effect of Fire A on the destruction of the lodge, and to take a moment to articulate why the traditional but-for analysis (i.e., conditioning on Fire B) leads to an incorrect conclusion—that is, to articulate why it is intuitively incorrect to conclude that Fire A is not a cause of the lodge’s destruction.

Many people, including myself, would explain that this conclusion is counter to our intuition, as well as court decisions, because it is known that, had Fire B not occurred, Fire A would have independently caused the destruction of the lodge (and, had Fire A not occurred, Fire B would have independently caused the destruction of the lodge). Therefore, the reasoning goes, it is illogical to conclude that neither fire is a cause of the outcome simply because the fires occurred simultaneously. As the U.S. Supreme Court has stated in a distinct but closely related context, “it would be nonsensical

101. See generally *Mitchell v. Gonzales*, 819 P.2d 872, 876 (Cal. 1991) (“It has generally been recognized that the ‘but for’ test . . . should not be used when two ‘causes concur to bring about an event and either one of them operating alone could have been sufficient to cause the result The proper rule for such situations is that the defendant’s conduct is a cause of the event because it is a material element and a substantial factor in bringing it about.” (quoting *Vecchione v. Carlin*, 168 Cal. Rptr. 571, 576 (Ct. App. 1980))); *DOBBS ET AL.*, *supra* note 1, at 322.

to adopt a rule whereby individuals hurt by the combined wrongful acts of many . . . would have no redress, whereas individuals hurt by the acts of one person alone would have a remedy.”¹⁰² Thus, we reason that, because either fire would have destroyed the lodge independently, both fires are causes of the lodge’s destruction.

But this logic is precisely the reasoning used in our main-effects analysis. By examining, for example, the main effect of Fire A in the potential outcomes framework, we ask not only what the outcome would be had Fire A not occurred but what the outcome would be had Fire B not occurred. Intuitively, we (and the courts) conclude that Fire A is a cause, because, had Fire B not occurred, Fire A would have destroyed the lodge. In our main-effects analysis, we conclude similarly: Fire A would have destroyed the lodge had Fire A been “on” and Fire B been “off,” thus affecting the proportion of combinations of Fire A and Fire B for which the lodge is destroyed when Fire A is “on” versus “off,” thereby leading to the conclusion that Fire A is a cause of the lodge’s destruction.

The point is this: In considering whether Fire A constitutes a cause of the lodge’s destruction, our common perception of what constitutes a cause is based not only on a comparison of counterfactuals associated with Fire A, but also on a comparison of counterfactuals associated with Fire B. Fundamental to our reasoning for why Fire A constitutes a cause is the question, what would have happened had Fire B not occurred? Accordingly, to examine both sets of counterfactuals, we examine the main effect of Fire A. Furthermore, this analysis is perfectly consistent with the counterfactual model of causation. It simply broadens the causation inquiry to account for the complexity of a multi-factor problem, pursuant to the potential outcomes framework.

C. *Does the Factorial Approach Change the Meaning of Counterfactual Causation?*

A possible response to my analysis is that the factorial approach permits the intuitive result only by changing the meaning of counterfactual causation—in particular, by violating the “rule” of the counterfactual model that we consider the outcome in a hypothetical world in which we remove the condition, *and only the condition*, that is alleged to have caused the subject injury. This response, however, is grounded in an oversimplified understanding of the counterfactual model and but-for causation.

102. *Paroline v. United States*, 572 U.S. 434, 452 (2014).

As discussed *supra*, factual causation is intended to reflect actual, or scientific, cause and effect. This is the underlying purpose of the counterfactual inquiry—to determine, as a matter of fact, whether a defendant’s wrongful act caused the plaintiff’s injury. The counterfactual model of causation is not a proxy for scientific cause and effect—it *is* the conceptual basis of scientific cause and effect, as represented by the potential outcomes framework. While most situations entail a straightforward application of the but-for standard, this standard is only a specific measure within the broader counterfactual framework. But it is this broader framework that the law seeks to capture in the factual-causation inquiry; and the same principles from this broader framework that underlie the but-for standard in single-factor problems extend to establish main-effects estimands in multi-factor problems. These estimands are consistent with both common sense and scientific cause and effect under the counterfactual model.

It is incorrect to say that the above analysis violates the “rules” of the counterfactual model of cause and effect. Further, because this is the case, and because the law intends a model of causation that reflects scientific cause and effect, it is also incorrect to say that the main-effects analysis above is invalid on the grounds that it changes the “rules” of counterfactual *causation*.

D. *Extending the Analysis to Successive Multiple Sufficient Causes*

Treating both fires as interventions and examining both interaction effects and main effects allows a more comprehensive understanding of the effects of each fire on the destruction of the lodge. Applying the factorial approach, each fire in the concurrent two-fire problem can be understood as a cause of the lodge’s destruction. Similar reasoning applies to the first fire in the *successive* two-fire problem.

Imagine that Fire A and Fire B arrived at the site of the lodge from slightly different directions, and that Fire A arrived prior to Fire B, thus preempting the ability of Fire B to destroy the lodge. We are interested in knowing the effect of Fire A on the destruction of the lodge. To analyze the problem, we can simply examine the effect of Fire A as we did in the concurrent two-fire problem. Doing so similarly results in a positive main effect of Fire A on the lodge’s destruction and an interaction effect of zero. Alternatively, we can use variations of this approach to incorporate explicitly the time element featured in the successive two-fire problem—although traditional formulations of the but-for approach to this problem (i.e., conditioning on the

occurrence of Fire B) often do not do so.¹⁰³ In any event, a main-effects analysis generally leads to the conclusion that Fire A is a cause of the lodge's destruction.

Applying a factorial analysis to the successive two-fire problem is relatively straightforward for analyzing the main and interaction effects of Fire A. However, defining the main effects of Fire B in the successive two-fire problem is somewhat more complex. In particular, although we know from the successive two-fire scenario that Fire B arrives only after Fire A has already destroyed the lodge, examining the main effect of Fire B could lead to the conclusion that Fire B has a positive effect on the lodge's destruction—a misleading result, given that a fire cannot cause the destruction of a lodge that has already been destroyed.¹⁰⁴ This conclusion is, however, avoided if the problem is structured carefully using a “sequential” design that accounts for the successive nature of the fires in the successive two-fire problem.¹⁰⁵

In a multi-treatment experiment, a sequential design allows the examination of multiple treatments that occur in succession rather than concurrently—exactly the scenario in the successive two-fire problem.¹⁰⁶ This is a developing area of causal inference, but valid and well-defined within the potential outcomes framework. In a two-factor sequential experiment, the idea is to apply one treatment in period one, measuring the outcome variable at the end of period one; and then to apply the second treatment in period two, measuring the outcome variable at the end of period two while conditioning on the outcome of period one.¹⁰⁷ Therefore, the causal effect would be defined as a contrast between the potential outcomes when Fire B is “on” and when Fire B is “off,” given the outcome of period 1, the destruction of the lodge. This leads to the calculation $1 - 1 = 0$: the lodge at the end of period 2 is destroyed when Fire B is “on” and when Fire B is “off,” conditional on the lodge having been destroyed in period 1. This analysis leads to the intuitive result that Fire B has no causal effect on the destruction of the lodge in the successive two-fire problem.

103. In particular, although the fact pattern in the successive two-fire problem involves successive fires, the literature applying the but-for test to this problem frequently does not incorporate the time element in any explicit way in its analysis.

104. Under some theories of causation, it could be acceptable to define Fire B as a cause; but, consistent with the overwhelming consensus in the courts and literature, I assume that Fire B is understood as incapable of causing the destruction of a lodge that has previously been destroyed by Fire A.

105. See generally Iavor Bojinov & Neil Shephard, *Time Series Experiments and Causal Estimands: Exact Randomization Tests and Trading*, J. AM. STAT. ASS'N (forthcoming 2019), <https://doi.org/10.1080/01621459.2018.1527225> [<https://perma.cc/LV2X-6HSR>].

106. See *id.* at 1–5.

107. See *id.*

Perhaps more simply, care needs to be taken in specifying treatments and causal effects, the definitions of which include a time element. Defining an appropriate estimand generally requires an ordering in which a treatment occurs prior to an outcome. In the following Part, I argue that the factorial approach can be applied in practice by specifying an estimand based on the causal-set approach of the NESS test and the Restatement Third. Under the NESS test, “a condition contributed to some consequence if and only if it was necessary for the sufficiency of a set of existing antecedent conditions that was sufficient for the occurrence of the consequence.”¹⁰⁸ It thus requires that the tortious act be “a necessary element of the set of *actual antecedent conditions* that was sufficient for the injury,” a “necessary condition” of which is “that the injury not have occurred already as a result of other actual conditions outside the set.”¹⁰⁹

The issue of successive multiple sufficient causes again highlights a general theme of this article: The potential outcomes model provides a coherent framework for analyzing causal questions; but the researcher (or the court) is responsible for defining estimands that are appropriate under the circumstances of a case. The time element in the successive two-fire problem provides a good example: oversimplification can lead to a misleading result.

More generally, the flexibility enabled by the potential outcomes framework could arguably lead to a degree of uncertainty. But the capacity of the potential outcomes framework to integrate different estimands should not be viewed as a weakness. Indeed, it is widely understood as an important feature of the framework. The law should, however, be clear regarding the meaning of relevant estimands and, for example, the circumstances under which a multi-treatment analysis is appropriate.

V. IMPLICATIONS

A. *The Counterfactual Model as a General Theory of Causation*

The purported inability to “reduce[] [causation] to a single concept” has caused substantial problems in various areas of the law.¹¹⁰ As discussed in Part II, courts and scholars have treated cases involving multiple sufficient causes as requiring an exception to the counterfactual model of causation. This purported failure of the counterfactual model has produced considerable

108. Wright, *Grounds and Extent*, *supra* note 32, at 1441; Wright, *Once More into the Bramble Bush*, *supra* note 32, at 1102–03; *see also* Wright, *supra* note 6, at 1788–1803.

109. Wright, *supra* note 6, at 1794–95 (emphasis added).

110. LANDES & POSNER, *supra* note 48, at 228.

opposition to the model and calls to altogether abandon it as a theory of causation. My results in Part IV show, however, that these efforts, and even efforts to abandon the but-for measure of causation in particular, are built on faulty ground.

As explained above, the problem arises from an oversimplified understanding of the counterfactual model and but-for causation. Although for most causation inquiries, the results that would emerge from a narrow and simple understanding of this model are consistent with our intuition and a broader understanding based on scientific cause and effect, courts and scholars have frequently rejected the counterfactual model based on an oversimplified application of it to the exceptional and complex circumstances of concurrent and successive multiple sufficient causes. These circumstances, however, require additional nuance, as evidenced by scientific methods for addressing multiple factors based on the counterfactual model.

Predominant methods for proving cause and effect in the hard and social sciences are based on the potential outcomes framework. This framework, which has been applied in legal contexts to prove statistical cause and effect, also allows for a careful and rigorous application of the counterfactual model of causation in cases involving multiple sufficient causes. Following the potential outcomes framework strengthens the causation standard by melding it with scientific standards of cause and effect and making it consistent with standards applied to statistical evidence offered to *prove* factual causation. Furthermore, however, it strengthens the law's notion of causation conceptually, allowing application in a way that maintains its simplicity and intuitive results in simple cases, while providing a well-established and robust framework for analyzing more complex cases that is consistent with both our intuition and scientific cause and effect.

The factorial approach to multiple sufficient causes examines the effects of hypothesized causes as "treatments" in a multi-treatment, or factorial, setting within the potential outcomes framework. Applying this approach, and a main-effects analysis in particular (*see* Part IV), to the two-fire problems leads to the conclusion that each fire in the concurrent two-fire problem and the first fire in the successive two-fire problem can and should be understood, consistent with intuition and court decisions, as causes of the lodge's destruction in those problems. Indeed, it leads to acceptable results more generally, because it permits a logical method for analyzing causal problems.

This conclusion is consistent not only with intuition, but with the predominant scientific conception of cause and effect based on the counterfactual model. Therefore, it is incorrect to say that the counterfactual model fails as applied to cases involving multiple sufficient causes or that an

exception to the counterfactual model is needed in such cases. Moreover, it is incorrect to abandon the counterfactual model as a theory of causation on the grounds that it fails in these cases or that it is an incomplete account of what is meant by cause and effect based on multiple-sufficient-cause situations. To the contrary, one must only look to statistics and the sciences—undoubtedly fundamental to our common conception, as well as tort law’s notion, of cause and effect—for appropriately defining factual causation in these circumstances.

B. Counterfactual Causation Should be Used in Torts Cases Involving Multiple Sufficient Causes

In the preceding section, I discuss descriptive implications of my results—specifically, that it is incorrect to reject the counterfactual model of causation on grounds that it fails in multiple-sufficient-cause situations. In this section, I address the separate issue of how to analyze and decide cases involving multiple sufficient causes. I argue that my results disfavor the substantial-factor test and that courts should instead apply a standard based on the factorial approach, which, I show, can be simplified and applied in practice by using the causal-set approach of the NESS test and the Restatement Third. Indeed, the causal-set approach can be understood as an application of the counterfactual model within the meaning of scientific cause and effect.

Tort law seeks a factual causal inquiry that generates outcomes that reflect actual, or scientific, cause and effect. Fundamental to this aim is a standard of factual causation that 1) reflects actual cause and effect and 2) permits a principled approach to determining whether the standard has been satisfied. The counterfactual model, and the factorial approach in particular, fulfills both of these criteria; the substantial-factor test satisfies neither. Additionally, applying the causal-set approach in the NESS test and the Restatement Third would allow courts to apply a simple standard that is consistent with and supported by the factorial approach.

1. Rejecting the substantial-factor test in favor of the factorial approach.

The factorial approach reflects the scientific notion of cause and effect and a proper application of the counterfactual model to cases involving multiple sufficient causes. This is not the case with the substantial-factor test. The

substantial-factor test—the most common alternative to the but-for test¹¹¹—has no analytical connection to the meaning of actual, or scientific, cause and effect. Relatedly, as commentators and courts, including the Supreme Court, have highlighted, the substantial-factor test lacks a well-defined meaning in the first instance, giving rise to unpredictability and outcomes that derive from the whims of factfinders rather than a principled analysis.¹¹²

It is not even clear how the substantial-factor test would be integrated with predominant conceptions of cause and effect. For example, an analysis of whether an event constitutes a substantial factor would begin by defining an outcome variable and an event that is hypothesized to have impacted the outcome variable. If the outcome variable is binary, such as whether a lodge was destroyed or not or whether a person died or not, then the substantial-factor test seems to ask whether the event “contributed” to the outcome. But what does it mean to “contribute” to an outcome? Perhaps it means that some percentage value can be ascribed to the event in producing the outcome. But this notion is ill defined and quickly crumbles unless it is transformed to a more complex version of the counterfactual model of cause and effect. It is possible, for example, to formulate an effect in terms of a percentage value by using 1) the simple but-for test, but where the outcome variable is defined as the percentage change in the *likelihood* of a binary event occurring rather than a binary outcome variable; or 2) the factorial approach’s main-effects analysis, which, like the but-for test, relies on the counterfactual model, but which employs an outcome variable that can be defined in terms of percentage values based on a comparison of potential outcomes associated with combinations of hypothesized causes. But, if the substantial-factor test is intended to refer to either of these notions, then let courts define it rigorously as such. The problem is that the substantial-factor test does not refer to either of these concepts; rather, it refers to a nebulous idea that converts the factual-cause inquiry to a standardless determination that reflects nothing but a vague sense of attributing responsibility for an outcome to some

111. See, e.g., *Rutherford v. Owens-Ill., Inc.*, 941 P.2d 1203, 1214 (Cal. 1997) (“California has definitively adopted the substantial factor test . . . which subsumes the ‘but for’ test while reaching beyond it to satisfactorily address other situations, such as those involving independent or concurrent causes in fact.”).

112. See, e.g., *Burrage v. United States*, 571 U.S. 204, 217 (2014) (commenting that the prosecution’s proposed use of a substantial-factor test “cannot be reconciled with sound policy, given the need for clarity and certainty in the criminal law” (internal quotation marks omitted)); *Rutherford*, 941 P.2d at 1214 (“The term ‘substantial factor’ has not been judicially defined with specificity, and indeed it has been observed that it is ‘neither possible nor desirable to reduce it to any lower terms’” (quoting KEETON ET AL., *supra* note 1, § 41, at 267); *Ford Motor Co. v. Boomer*, 736 S.E.2d 724, 730 (Va. 2013) (“In sum, some jurors might construe the term to lower the threshold of proof required for causation while others might interpret it to mean the opposite. We do not believe that substantial contributing factor has a single, common-sense meaning . . .”).

event. Consequently, courts employing this standard leave the liability determination to a jury's intuitive sense of responsibility rather than a principled factual inquiry, resulting in a high degree of randomness and bias through variables that would not be relevant to a factual inquiry regarding cause and effect.¹¹³

The factorial approach is more principled. In the two-fire scenario, a factfinder could complete the 2x2 table containing outcomes associated with combinations of Fire A = on, Fire A = off, Fire B = on, and Fire B = off, as illustrated in Figure 2.

	Fire A = off	Fire A = on
Fire B = off	?	?
Fire B = on	?	?

Figure 2. The factorial approach requires completing the 2x2 table containing potential outcomes associated with combinations of Fire A and Fire B.

A multiple-sufficient-cause situation arises from findings of fact consistent with those in the 2x2 table in Figure 3. Under these circumstances, Fire A and Fire B can be understood as causes based on a main-effects analysis reflecting these findings—for example, based on the fact that Fire A would destroy the lodge in a greater proportion of treatment combinations when Fire A is on than when Fire A is off; and Fire B would destroy the lodge in a greater proportion of treatment combinations when Fire B is on than when Fire B is off. Using this approach, to determine whether Fire A is a cause of the lodge's destruction, a factfinder would complete the table and answer the question: "What is the impact of Fire A being on versus off on the proportion of combinations for which the lodge is destroyed?"

113. "[T]he development of several quite distinct and conflicting meanings for the term 'substantial factor' has created risk of confusion and misunderstanding, especially when a court, or an advocate or scholar, uses the phrase without explicit indication of which of its conflicting meanings is intended." *Mitchell v. Gonzales*, 819 P.2d 872, 884 (Cal. 1991) (Kennard, J., dissenting) (quoting *KEETON ET AL.*, *supra* note 1, 1988 supp., § 41, at 43) (internal quotation marks omitted); see also Gerald W. Boston, *Toxic Apportionment: A Causation and Risk Contribution Model*, 25 ENVTL. L. REV. 549, 631 (1995) (describing the substantial-factor test as one that "has become a default, resorted to when nothing else works" and for which "juries are afforded virtually no guidance as to how much of a causal connection is necessary to satisfy the test").

The factorial approach is thus relatively simple and applies generally to multiple-sufficient-cause situations, including more complex problems such as those involving more than two factors. I discuss such problems and implications of the factorial approach for the causal-set model in the following subsection.

Multiple Sufficient Causes		
	Fire A = off	Fire A = on
Fire B = off	Not Destroyed	Destroyed
Fire B = on	Destroyed	Destroyed

Figure 3. 2x2 table illustrating potential outcomes in a multiple-sufficient-cause situation.

2. The causal-set approach.

The factorial approach is a rigorous method, based on the counterfactual model, for analyzing actual cause and effect. Using certain modest assumptions, it can generally be applied in practice—with similar logic and results—by using the causal-set approach of the NESS test and the Restatement Third. Let us compare the Restatement Third approach to the factorial approach in the concurrent two-fire problem.

Section 27 of the Restatement Third states: “If multiple acts occur, each of which under § 26 alone would have been a factual cause of the physical harm at the same time in the absence of the other act(s), each act is regarded as a factual cause of the harm.”¹¹⁴ This standard requires the same findings of fact as those required in the factorial approach and as illustrated in the 2x2 matrices in Figures 2 and 3. Section 27 implicitly requires that the occurrence of both Fires together is associated with the destruction of the lodge (lower right quadrant in Figure 3) and explicitly requires that 1) the occurrence of Fire A alone is associated with the destruction of the lodge (upper right quadrant of Figure 3), 2) the occurrence of Fire B alone is associated with the destruction of the lodge (lower left quadrant of Figure 3), and 3) the non-occurrence of Fire A and Fire B is associated with the non-destruction of the lodge (upper left quadrant of Figure 3). Therefore, making certain modest assumptions regarding the estimand to be applied in the factorial approach, the Restatement Third standard (as well as the NESS test, a version of which

114. RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL & EMOTIONAL HARM § 27 (AM. LAW INST. 2010).

is incorporated in the Restatement Third approach) applies the same basic algorithm and produces the same results as does the factorial approach in the concurrent two-fire scenario and analogous problems: both methods complete the matrix in Figure 2 and both result in the finding that Fire A and Fire B are causes if the findings of fact lead to the pattern in Figure 3.¹¹⁵

Thus, the factorial approach can be understood as supporting—and specifically, as providing a scientific foundation for—the causal-set approach in the Restatement Third (and the NESS test). This foundation could replace an outcome-oriented understanding of the Restatement Third approach—one, for example, that views this approach as “a legal principle to govern the *outcome*” directly and to “declare[.]” acts as factual causes where the but-for test, a test based on the actual or scientific meaning of cause and effect, fails and “further assistance” is needed.¹¹⁶

Let us now compare the NESS test and the factorial approach in situations involving more than two factors.¹¹⁷ Imagine an action involving allegations that five polluters caused a certain type of pollution to rise to dangerous levels, causing injury.¹¹⁸ Employing the factorial approach, computing the factorial main effects of Polluter A in particular would involve a comparison between the potential outcomes associated with Pollutant A = on and the potential outcomes associated with Pollutant A = off, aggregating over the various combinations of the occurrence and non-occurrence of the other pollutants. Using the basic notation of Part II, the problem would involve a comparison between 1) the potential outcomes $Y_1(D_a=1, D_b, D_c, D_d, D_e)$, $Y_2(D_a=1, D_b, D_c, D_d, D_e)$, . . . , $Y_m(D_a=1, D_b, D_c, D_d, D_e)$, where D_a, \dots, D_e represent indicators for whether each of the five factors are on or off, $D_a=1$ indicates that Pollutant A = on, and subscripts $1, \dots, m$ index the particular treatment combination of $D_a=1, \dots, D_e$ with which each potential outcome is associated, and 2) the potential outcomes $Y_{m+1}(D_a=0, D_b, D_c, D_d, D_e)$, $Y_{m+2}(D_a=0, D_b, D_c, D_d, D_e)$, . . . , $Y_n(D_a=0, D_b, D_c, D_d, D_e)$, where $D_a=0$ indicates that Pollutant A = off and subscripts $m+1, \dots, n$ index the particular treatment

115. The causal-set approach, like the traditional but-for measure of causation, exists within the potential outcomes framework—it defines certain measures within the broader counterfactual model.

116. *State v. Tribble*, 790 N.W.2d 121, 127 (Iowa 2010) (emphasis added).

117. *See, e.g., Major v. R.J. Reynolds Tobacco Co.*, 222 Cal. Rptr. 3d 563, 581 (Ct. App. 2017) (involving multiple tobacco producers and asbestos manufacturers; “the issue raised . . . is how to deal with a case of *more than two* concurrent causes, when various combinations of the causes—although perhaps not any individual cause—would have been sufficient to cause the harm”).

118. *See generally* RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYSICAL & EMOTIONAL HARM § 27 cmt. f, illus. 3 (AM. LAW INST. 2010).

combination of $D_a=0, \dots, D_e$ with which each potential outcome is associated.

In essence, the factorial main-effects inquiry asks whether dangerous levels of pollution occur for a higher proportion of treatment combinations in which Pollutant A = on than treatment combinations in which Pollutant A = off. Frequently—and specifically, when the number of treatment combinations in the two categories are equal—this is equivalent to the question whether dangerous levels of pollution occur more frequently when Pollutant A = on or Pollutant A = off. However, to avoid considering all of the treatment combinations in each category, we can simply ask whether, for a particular treatment combination, Pollutant A would mean the difference between dangerous levels of pollution and non-dangerous levels of pollution. For example, if the potential outcomes associated with two treatment combinations, the only difference between which is that Pollutant A is on versus off, are different—and specifically, reaching the pollution threshold for Pollutant A = on and not reaching it for Pollutant A = off—then, holding all else equal, this can be interpreted as a causal effect of Pollutant A in a main-effects analysis. Notationally, if $Y_i(1, d_b, d_c, d_d, d_e) - Y_j(0, d_b, d_c, d_d, d_e) = 1$ (otherwise stated, if $Y_i(1, d_b, d_c, d_d, d_e) = 1$ and $Y_j(0, d_b, d_c, d_d, d_e) = 0$), where lowercase d_a, \dots, d_e are used to represent particular values of D_a, \dots, D_e , respectively, and i and j represent particular treatment combinations, then, putting aside other possible differences in pairs of treatment combinations when Pollutant A = on versus Pollutant A = off, Pollutant A can be said to have a factorial main effect on reaching dangerous levels of pollution.¹¹⁹

Significantly, however, this simplified formulation of the factorial main-effects approach constitutes a form of the causal-set approach of the NESS test, which could provide a straightforward prescription for courts to apply in cases involving multiple hypothesized causes. The NESS test asks whether Pollutant A is a necessary element of a sufficient set—that is, whether it is necessary to the sufficiency of a set in producing a dangerous level of pollution. In the language of the potential outcomes framework, this test can be formulated as asking whether there is a pair of treatment combinations, the only difference between which is that one involves Pollutant A = on and the other involves Pollutant A = off, for which the potential outcome associated with Pollutant A = on is 1 (i.e., reaching the pollution threshold) and the potential outcome associated with Pollutant A = off is 0 (i.e., not reaching the pollution threshold). Notationally, this formulation asks whether there is a pair of potential outcomes such that

119. This logic in fact relies on the reasonable assumption that the emission of Pollutant A can add to the pollution and to the possibility of reaching the threshold but cannot diminish the pollution or the possibility of reaching the threshold.

$Y_i(1, d_b, d_c, d_d, d_e) - Y_j(0, d_b, d_c, d_d, d_e) = 1$. This is the same question that is asked above in the context of the simplified factorial main-effects analysis.

To be sure, there are important differences between the NESS test and the potential outcomes framework, and the factorial approach in particular. Nevertheless, the logic and the results of the NESS test will frequently mirror those of the factorial approach. The same can be said of the Restatement Third approach.¹²⁰

Therefore, as in the two-fire problem, in addition to providing a rigorous framework for addressing multiple sufficient causes using the counterfactual model, the factorial approach supports the causal-set model of the NESS test and the Restatement Third by providing a rigorous scientific and statistical foundation for them, and, further, by establishing their fundamental connection to the counterfactual model of cause and effect.

C. *Implications Beyond Tort Law*

Many areas of the law look to tort law to decide cases involving multiple sufficient causes and complex causation issues in general. This is, in part, due to the extensive analysis afforded these issues in tort law, as well as the close analogy between these issues in torts cases and other areas of the law.¹²¹ Such areas include criminal law, contracts, employment discrimination, and antitrust law, to name a few. For example, in the criminal case *Burrage v. United States*, the Supreme Court relied on tort law to determine whether a defendant's distribution of heroin constituted a cause of a drug user's death, where the drug user died following a binge that included the heroin but that may have been sufficient to kill the drug user even without the heroin.¹²² In the employment-discrimination context, the Supreme Court has held that "in

120. Philosophic components underlying NESS overlap substantially with those underlying the potential outcomes framework. The meaning of cause and effect in the potential outcomes framework, or the Rubin Causal Model, is in various respects consistent with the philosophic notions of cause and effect of Hume and Mill. Paul W. Holland, *Statistics and Causal Inference*, 81 J. AM. STAT. ASS'N 945, 951 (1986). As statistician Paul Holland concluded, "Mill's thinking, being driven by an experimental model, is in reasonably close agreement with the [Rubin Causal Model]." *Id.* At the same time, the NESS test "incorporates the traditional Humean philosophic account of the meaning of causation, as modified by John Stuart Mill." Wright, *supra* note 6, at 1774.

121. See *Univ. of Tex. Sw. Med. Ctr. v. Nassar*, 570 U.S. 338, 342 (2013) ("The requisite relation between prohibited conduct and compensable injury is governed by the principles of causation, a subject most often arising in elaborating the law of torts.").

122. See *Burrage v. United States*, 571 U.S. 204, 210–17 (2014); see also *Paroline v. United States*, 572 U.S. 434, 452 (2014); *State v. Tribble*, 790 N.W.2d 121, 126–27 (Iowa 2010) (relying on tort law in felony-murder case to define causation generally and in multiple-sufficient-cause situations).

enacting Title VII,” Congress “is presumed to have incorporated” the cause-in-fact standard of tort law “absent an indication to the contrary in the statute itself.”¹²³ And, in antitrust law, courts frequently refer to tort law in resolving complex causation issues.¹²⁴

In these contexts, and in others, the meaning of causation and the operation of the counterfactual model are tightly connected to those in tort law, either directly—because the law in these substantive areas relies on tort law—or at least by analogy. As such, the implications for tort law discussed *supra* generally extend to issues of causation, and of multiple sufficient causes in particular, in these substantive areas.

Consider, for example, a circumstance in contract law that can give rise to multiple-sufficient-cause issues analogous to those in tort law: In *California & Hawaiian Sugar Co. v. Sun Ship, Inc.*, a plaintiff agriculture cooperative contracted with two companies for the production of an “integrated tug barge” (one company for the construction of a barge and the second company for the delivery of the tug) to transport raw sugar by sea from Hawaii to California.¹²⁵ Both companies breached their respective contracts by delivering on them late, causing substantial injury.¹²⁶ The Ninth Circuit held that “in this case of concurrent causation each defaulting contractor is liable for the breach and for the substantial damages which the joint breach occasions.”¹²⁷ Applying a form of the substantial-factor test, the court held that the defendant was “a substantial cause of the damages flowing from the lack of the integrated tug; [it] cannot be absolved by the absence of the tug.”¹²⁸

My conclusions above apply: rather than resigning simply to intuition and applying the nebulous substantial-factor test based on the purported failure of the counterfactual model in this scenario, the factorial approach may allow

123. *Nassar*, 570 U.S. at 346–47; see also *Meyer v. Holley*, 537 U.S. 280, 285 (2003) (“This Court has noted that an action brought for compensation by a victim of housing discrimination is, in effect, a tort action. . . . And the Court has assumed that, when Congress creates a tort action, it legislates against a legal background of ordinary tort-related vicarious liability rules and consequently intends its legislation to incorporate those rules.”).

124. See, e.g., *Lotes Co. v. Hon Hai Precision Indus. Co.*, 753 F.3d 395, 415 (2d Cir. 2014) (remarking, in case brought under Sections 1 and 2 of the Sherman Act, that case does not involve circumstances in which “an injurious event is ‘overdetermined’ by multiple sufficient causes,” citing RESTATEMENT (THIRD) OF TORTS: PHYSICAL & EMOTIONAL HARM § 27 (AM. LAW INST. 2010)); *Jack Walters & Sons Corp. v. Morton Bldg., Inc.*, 737 F.2d 698, 708–09 (7th Cir. 1984) (commenting that a particular antitrust doctrine is simply “the application to antitrust law of venerable principles of tort causation”); *Valley Prods. Co. v. Landmark*, 877 F. Supp. 1087, 1092 (W.D. Tenn. 1994) (“[T]he [Supreme] Court recognized a substantive causation element in antitrust claims, analogous to the notion of proximate cause in tort actions.”).

125. 794 F.2d 1433, 1434–35 (9th Cir. 1986).

126. *Id.* at 1435.

127. *Id.* at 1437.

128. *Id.* at 1437–38.

a rigorous test that follows the counterfactual model. More generally, the factorial approach would facilitate a better standard for addressing multiple-sufficient-cause situations in a wide range of legal contexts that directly follow or are analogous to these situations and accompanying law in torts cases.

D. *The “Ordinary Meaning” of Causal Language*

Another general context for which the analysis in Part IV has important implications involves the common situation in which a case turns on judicial interpretation of the “ordinary meaning” of causal language in a statute, such as “results from,” “because of,” or “based on.” When a statute does not specify the meaning of a phrase, courts interpret it using the phrase’s “ordinary meaning.”¹²⁹ The Supreme Court has held that phrases such as “results from,” “because of,” and “based on” are “ordinarily” defined as but-for causation.¹³⁰ In *Burrage*, for example, the Court interpreted the Controlled Substances Act to determine whether a drug user’s death “following an extended drug binge” that involved, among other drugs, heroin purchased from the defendant fell within the Act’s language imposing a minimum sentence if the death “resulted from” the defendant’s unlawful distribution.¹³¹ According to the Court:

The Controlled Substances Act does not define the phrase “results from,” so we give it its ordinary meaning. See *Asgrow Seed Co. v. Winterboer*, 513 U.S. 179, 187, 115 S.Ct. 788, 130 L.Ed.2d 682 (1995). A thing “results” when it “[a]rise[s] as an effect, issue, or outcome *from* some action, process or design.” 2 *The New Shorter Oxford English Dictionary* 2570 (1993). “Results from” imposes, in other words, a requirement of actual causality. “In the usual course,” this requires proof “‘that the harm would not have occurred’ in the absence of—that is, but for—the defendant’s conduct.” *University of Tex. Southwestern Medical Center v. Nassar*, 570 U.S. —, —, 133 S.Ct. 2517, 2525, 186 L.Ed.2d 503 (2013) (quoting Restatement of Torts § 431, Comment *a* (1934)).¹³²

The Court held that “[t]his but-for requirement is part of the common understanding of cause,” and noted that “courts regularly read phrases like

129. *Burrage v. United States*, 571 U.S. 204, 210 (2014).

130. *Id.* at 210–14.

131. *Id.* at 206–07.

132. *Id.* at 210–11.

‘results from’ to require but-for causality.’¹³³ In *Burrage*, the Court relied on this ordinary meaning of “results from” to hold that the prosecution did not satisfy the causation element of its claim. The Court rejected the prosecution’s proposed “less demanding” “substantial” or “contributing” factor test as harmful to goals of “clarity and certainty in the criminal law.”¹³⁴ It held, however, that it was not necessary to decide on how the statute’s language would apply to multiple-sufficient-cause situations, since there was no evidence that the distributed heroin “was an independently sufficient cause of [the decedent’s] death.”¹³⁵

Although the Court expressly declined to indicate how its analysis would apply to multiple-sufficient-cause situations, its discussion highlights certain complexities that would arise in these circumstances, including the Court’s reluctance to apply a substantial-factor test and its interest in adhering to the ordinary meaning of the statute’s “results from” language. In these circumstances, the Court’s analysis could arguably benefit from the factorial approach, which would provide the Court with an avenue to adhere to the ordinary meaning of the statute’s language (and policy reasoning suggested in *Burrage*) by applying the counterfactual model of causation and avoiding a vague substantial-factor test, while at the same time allowing the Court to uphold the outcome in multiple-sufficient-cause situations that courts have near-universally determined (and that statutes such as the Controlled Substances Act may intend)—that a concurrent sufficient cause constitutes a “cause.”

In particular, the Court could determine that, in light of the meaning of scientific cause and effect as represented by the potential outcomes framework, the “ordinary” meaning of causation is broader than the traditionally narrow understanding of but-for causation and, rather, entails the broader counterfactual model and the possibility of measures based on main effects as well as interaction effects. This interpretation is consistent with both approaches discussed in section III.B—broadening the scope of but-for causation to incorporate principles from the broader counterfactual model or looking directly to the broader counterfactual model for the meaning of causation in multiple-sufficient-cause situations. Further, this interpretation is supported by the argument in section III.B that the “ordinary” meaning of causation, as reflected in standard dictionary entries for “causation” (and similar terms), is very closely related to, if not the same as, the scientific meaning of cause and effect.

133. *Id.* at 211–12.

134. *Id.* at 215, 217.

135. *Id.* at 214–15.

E. Implications for “Mixed-Motive” Employment Discrimination Cases

Finally, my analysis carries important implications for the employment discrimination context, and mixed-motive situations in particular. This category of applications can perhaps be understood as a subset of those discussed in the previous two sections; but the problems that arise from this context are sufficiently unique and important to merit an independent discussion.

A mixed-motive case is a type of employment discrimination case that involves an “employment decision attributable to . . . both legitimate and illegitimate factors.”¹³⁶ These cases arise frequently and have caused substantial inconsistency and confusion in employment discrimination law. This is not surprising: Although often not recognized as such, many mixed-motive situations constitute a specific type of multiple-sufficient-cause situation.¹³⁷ This is because, generally, these cases involve a situation in which two (or more) independent factors produce an outcome—an adverse employment action—where each alone would be sufficient to produce this outcome.

The Supreme Court case *University of Texas Southwestern Medical Center v. Nassar* involved a Title VII retaliation claim (among other claims) asserting that University of Texas Southwestern Medical Center retaliated against an employee physician, resulting in his firing, after he accused his superior of discrimination based on race and religion.¹³⁸ The Fifth Circuit upheld the retaliation claim based on the theory that these claims “require only a showing that retaliation was a motivating factor for the adverse employment action, rather than its but-for cause,” and its holding that “the evidence supported a finding that [the adverse employment action] was motivated, at least in part, to retaliate against respondent for his complaints against [his superior].”¹³⁹

136. *Univ. of Tex. Sw. Med. Ctr. v. Nassar*, 570 U.S. 338, 370 (2013) (Ginsburg, J., dissenting).

137. See generally Andrew Verstein, *The Failure of Mixed-Motives Jurisprudence*, 86 U. CHI. L. REV. 725, 728, 741–62 (2019) (arguing against application of the but-for standard in mixed-motive cases, concluding that “widespread acceptance of the But-For standard represents a great failure of the jurisprudence of mixed motives,” *id.* at 728, and that “many motive inquiries are not causal,” *id.* at 742, and, even “[i]f we analogize a defendant’s motives as potential causes of the action, then mixed motives are analogous to torts with multiple causes,” *id.* at 755; highlighting that “an item failing the but-for test can still [in tort law and in general] be a cause,” such as in cases involving multiple sufficient causes, *id.* at 754–55).

138. *Nassar*, 570 U.S. at 344–45 (majority opinion).

139. *Id.* at 345–46.

The Supreme Court, however, reversed the decision of the Fifth Circuit. Expressly applying tort law, it held that the but-for standard applies to determine whether the causation element was fulfilled:

It is thus textbook tort law that an action “is not regarded as a cause of an event if the particular event would have occurred without it.” This, then, is the background against which Congress legislated in enacting Title VII, and these are the default rules it is presumed to have incorporated, absent an indication to the contrary in the statute itself.¹⁴⁰

To provide some important background: In 1989, six justices of the Supreme Court agreed in *Price Waterhouse v. Hopkins* (although not in a majority opinion) that, to satisfy the causation element in the statutory language prohibiting discrimination against an individual “because of such individual’s race, color, religion, sex, or national origin,” a plaintiff need only “show that one of the prohibited traits was a ‘motivating’ or ‘substantial’ factor in the employer’s decision.”¹⁴¹ If this standard is satisfied,

the burden of persuasion would shift to the employer, which could escape liability if it could prove that it would have taken the same employment action in the absence of all discriminatory animus. In other words, the employer had to show that a discriminatory motive was not the but-for cause of the adverse employment action.¹⁴²

Congress then passed the Civil Rights Act of 1991, which “codified the burden-shifting and lessened-causation framework of *Price Waterhouse* in part but also rejected it to a substantial degree.”¹⁴³ Specifically, Congress amended Title VII to incorporate the requirement that a plaintiff show that “race, color, religion, sex, or national origin was a *motivating factor* for any employment practice, even though other factors also motivated the practice”;¹⁴⁴ at the same time, Congress replaced the Court’s burden-shifting scheme with one in which a showing by the employer that it would have taken the same employment action even without the influence of the plaintiff’s trait would nevertheless permit the plaintiff to win “declaratory relief, attorney’s fees and costs, and some forms of injunctive relief,” but not “monetary damages [or] a reinstatement order.”¹⁴⁵

140. *Id.* at 347 (citation omitted) (quoting KEETON ET AL., *supra* note 1, at 265).

141. *Id.* at 348 (quoting 42 U.S.C. § 2000e-2(a)) (citing *Price Waterhouse v. Hopkins*, 490 U.S. 228, 258 (1989) (plurality opinion)).

142. *Id.* (citations omitted) (citing *Price Waterhouse*, 490 U.S. at 258, 259–60, 276–77).

143. *Id.*

144. 42 U.S.C. § 2000e-2(m) (emphasis added).

145. *Nassar*, 570 U.S. at 349.

The Court again addressed this issue in *Gross v. FBL Financial Services, Inc.* in the context of a claim based on the Age Discrimination in Employment Act (ADEA), a statute establishing liability for an employer for discrimination against an individual “because of such individual’s age.”¹⁴⁶ In that case, the Supreme Court held that the ordinary meaning of the words “because of” is “by reason of: on account of,” and that the language, “because of . . . age” means “that age was the ‘reason’ that the employer decided to act”—i.e., “that age was the ‘but-for’ cause of the employer’s adverse decision.”¹⁴⁷ The Court rejected the argument that the Court’s interpretation of the statute is subject to the burden-shifting scheme established in *Price Waterhouse* for mixed-motive Title VII cases, “[n]oting that the ADEA must be read . . . the way Congress wrote it,” and citing “textual differences between Title VII and the ADEA” and “congressional choice not to add a provision like § 2000e-2(m) to the ADEA despite making numerous other changes to the latter statute in the 1991 Act.”¹⁴⁸

In *Nassar*, the Supreme Court addressed Title VII’s antiretaliation provision in particular. The statute provides that it is unlawful for an employer to discriminate against an employee “because he has opposed any practice made an unlawful employment practice by this subchapter, or because he has made a charge, testified, assisted, or participated in any manner in an investigation, proceeding, or hearing under this subchapter.”¹⁴⁹ The Court in *Nassar*, again focusing on the meaning of the term “because,” held that “[g]iven the lack of any meaningful textual difference between the text in this statute and the one in *Gross*, the proper conclusion here, as in *Gross*, is that Title VII retaliation claims require proof that the desire to retaliate was the but-for cause of the challenged employment action.”¹⁵⁰

Currently, therefore, the law provides for one causation standard, the motivating-factor standard, for some forms of discrimination, such as discrimination based on race, color, religion, sex, or national origin under Title VII, and a different causation standard for other forms of discrimination, such as discrimination based on age under the ADEA, as well as for retaliation under Title VII.

Moreover, even more muddled is the law surrounding mixed-motive cases based on the American with Disabilities Act (ADA), a statute prohibiting

146. 29 U.S.C. § 623(a); *Nassar*, 570 U.S. at 349–50.

147. *Gross v. FBL Fin. Servs., Inc.*, 557 U.S. 167, 176 (2009) (quoting 1 WEBSTER’S THIRD NEW INTERNATIONAL DICTIONARY 194 (1966)).

148. *Nassar*, 570 U.S. at 350–51 (internal quotation marks and citations omitted).

149. § 2000e-3(a).

150. *Nassar*, 570 U.S. at 352.

discrimination “on the basis of” disability.¹⁵¹ In this context, circuit courts are split, with some courts holding that the ADA’s causal language must be interpreted to require that a plaintiff’s disability be the sole cause of an adverse employment action, and other courts holding that a plaintiff’s disability need only be a motivating factor of the employer’s adverse employment action rather than the sole cause of it.¹⁵² In this area of employment discrimination law, as in others, there is confusion and uncertainty regarding the law within circuits as well as between them.¹⁵³

These contradictory, uncertain, and often vague standards governing mixed-motive claims grounded in Title VII, the ADEA, the ADA, and other discrimination statutes are based on statutory interpretation. The Supreme Court and lower courts have, however, frequently taken for granted the meaning of causal language in these statutes, such as “because of” and “on the basis of.” In cases discussed above, the Supreme Court has concluded that the “ordinary meaning” of such language implies a but-for standard, and that applying a but-for standard to mixed-motive cases involving a sufficient legitimate motivating factor necessarily precludes the conclusion that the illegitimate motivating factor is a “cause” of the adverse employment action. However, based on the scientific meaning of cause and effect, as well as the close relationship between the scientific and “ordinary” meanings of cause and effect, the causal language in these statutes imply a *counterfactual* model of causation but not necessarily a *but-for* standard. And, to the extent that such language can be said to imply a but-for standard, that standard can be interpreted as capturing main effects as well as interaction effects in the context of multiple-sufficient-cause situations, including mixed-motive cases.

151. 42 U.S.C. § 12112(a).

152. *Compare, e.g.,* Parker v. Sony Pictures Entm’t, Inc., 260 F.3d 100, 105, 107–08 (2d Cir. 2001) (reiterating an earlier ruling that “a plaintiff could succeed on his ADA claim where his disability played a motivating role in the decision, even if another factor . . . also played a substantial role” (internal quotation marks and citations omitted)), *with* Serwatka v. Rockwell Automation, Inc., 591 F.3d 957, 961–64 (7th Cir. 2010) (citing *Gross* and holding that “given the lack of a provision in the ADA recognizing mixed-motive claims, such claims do not entitle a plaintiff to relief for disability discrimination”), *and* Gulliford v. Schilli Transp. Servs., Inc., No. 4:15-CV-19-PRC, 2017 WL 1547301, at *6 (N.D. Ind. Apr. 27, 2017) (noting the 2008 amendments to the ADA changing the “because of” language relied on in *Serwatka* to “on the basis of,” but following the decision in *Serwatka* based on the finding “that there is no meaningful textual difference between ‘on the basis of’ and the term[] ‘because of’”). *See also* BARBARA T. LINDEMANN ET AL., EMPLOYMENT DISCRIMINATION LAW 197–99 (5th ed. 2012) (discussing circuit split and citing cases).

153. *See* Silk v. Bd. of Trs., Moraine Valley Cmty. Coll., Dist. No. 524, 795 F.3d 698, 705–06 (7th Cir. 2015) (describing tortuous history and present uncertainty surrounding law governing ADA claims based on mixed-motive theories).

Whether the but-for standard is interpreted narrowly or broadly is largely semantic. Either way, the substantive point remains the same: the ordinary meaning of causal language such as “because of” and “on the basis of” should be understood as capturing the broader counterfactual model and the possibility of causal measures based on main effects as well as interaction effects. In a single-factor situation, this meaning simplifies to the traditional narrow but-for standard. But, in multiple-sufficient-cause situations, the scientific and common meaning of causation entails a broader range of possibilities; and, furthermore, as suggested by the discussion in Part IV, there are good arguments for employing a standard of causation that reflects a main-effects analysis in particular.¹⁵⁴

Therefore, the ordinary meaning of the causal language in Title VII, the ADEA, and the ADA does not necessitate the conclusion that Congress intended to exclude illegitimate underlying purposes from the scope of unlawful activity in mixed-motive cases. Rather, a factorial approach to these cases implies that a motivating factor could be a “cause,” even when there are other sufficient motivating factors. Furthermore, the Court’s conclusion in *Gross* and *Nassar* that causal language such as the term “because of” implies a but-for standard does not, based on the counterfactual model, necessitate the further conclusion that Congress intended to preclude mixed-motive discrimination claims. Therefore, based on this reasoning, the Court could proceed to determine directly (rather than, e.g., through the term “because of”) whether Congress intended a causal relationship in the sense of main effects or interaction effects¹⁵⁵—that is, whether Congress intended for an illegitimate motivating factor among sufficient legitimate motivating factors to constitute a cause.

Without taking a position on whether Congress intended for such motivating factors to be captured by these statutes, I note that there are substantial arguments in favor of such a construction. For example, if a court cannot rely on the ordinary meaning of “because of” to exclude illegitimate motives in mixed-motive cases from unlawful behavior, precedent from torts and other areas of the law in which multiple sufficient causes are deemed “causes,” as well as other areas in the employment discrimination context, may play a more prominent role in interpreting congressional intent.

Furthermore, it is important to emphasize that the “motivating-factor” test is a particular form of the substantial-factor test described above. All of the criticisms that apply to the latter also apply to the former, including concerns

154. As in section V.D, this argument is supported by the analysis in section III.B concluding that the “ordinary” meaning of causation correlates extremely well with, if it is not altogether equivalent to, the scientific meaning of cause and effect.

155. *See supra* Part IV.

regarding the need for a clear standard of causation that is not simply left to the intuition of the factfinder.¹⁵⁶ As explained above, the factorial approach offers a more rigorous and well-defined standard.

VI. CONCLUSION

Courts should look to the fields of statistics and the sciences for guidance in developing a suitable theory of causation, as they very frequently do for evidence of causation. In particular, courts should adopt the potential outcomes framework for guidance in addressing causation inquiries in multiple-sufficient-cause situations.

Currently, courts and scholars treat multiple-sufficient-cause situations as a single-factor problem in which counterfactuals associated with a single variable—for example, Fire A in the two-fire problem—are compared while holding all other features of the problem, including other “sufficient causes,” constant. But this is contrary to how courts, scholars, and people generally *think* about the problem. Rather, we interpret both fires in the two-fire problem as “treatments,” or as “factors” in a factorial experiment: we reason that Fire A is a cause of the lodge’s destruction because, *had Fire B not existed*, Fire A would have destroyed the lodge, whereas (still assuming that Fire B had not existed) the lodge would not have been destroyed had Fire A not existed. That is, intuitively, we treat Fire A *and* Fire B as treatments and consider not only counterfactuals associated with Fire A, but also counterfactuals associated with Fire B, as suggested by the italicized language above.

Moreover, this reasoning is perfectly consistent with scientific cause and effect under the potential outcomes framework. That is, the counterfactual model itself does not require that we analyze causation as a single-factor problem. To the contrary, applying guidance from the potential outcomes

156. Note that the issue of whether a court should apply a but-for standard or a substantial-or motivating-factor standard of causation in cases involving multiple sufficient causes, some legitimate and some illegitimate, arises in other contexts also. For example, in *Mt. Healthy City Sch. Dist. Bd. of Educ. v. Doyle*, the Supreme Court, evaluating whether a school district’s refusal to renew the contract of a teacher violated his First and Fourteenth Amendment rights, held that “[a] rule of causation which focuses solely on whether protected conduct played a part, ‘substantial’ or otherwise, in a decision not to rehire, could place an employee in a better position as a result of the exercise of constitutionally protected conduct than he would have occupied had he done nothing”; and that, although “the burden was properly placed upon respondent to show that his conduct was constitutionally protected, and that this conduct was a ‘substantial factor’ or to put it in other words, that it was a ‘motivating factor’ in the Board’s decision not to rehire him,” “the District Court should have gone on to determine whether the Board had shown by a preponderance of the evidence that it would have reached the same decision as to respondent’s reemployment even in the absence of the protected conduct.” 429 U.S. 274, 285, 287 (1977).

framework, the problem could and should be examined as a multi-factor problem in which we consider two sets of counterfactuals, those associated with Fire A and those associated with Fire B.

Treating a multiple-sufficient-cause situation rigidly as a single-factor problem—contrary to scientific treatment of analogous problems and indeed contrary to our intuitive interpretation of the situation—it is not surprising that we arrive at a counterintuitive result.

Applying the potential outcomes framework, and the proposed factorial approach in particular, to cases involving multiple sufficient causes, we arrive at the intuitive result: both fires in the concurrent two-fire problem and the first fire in the successive two-fire problem should be understood as causes of the lodge's destruction.

This result carries important implications. First, courts and scholars have rejected the counterfactual theory of causation on grounds that it fails in circumstances involving multiple sufficient causes. But it does not; we have simply failed in applying it. Therefore, it is incorrect to apply this purported failure as the basis for a general move away from the counterfactual model. Moreover, courts have unjustifiably abandoned counterfactual causation in favor of the vague and intuition-based substantial-factor test in cases involving multiple sufficient causes. Courts have used the substantial-factor test as a policy tool, simply to replace a factual causal inquiry with an approach aimed at obtaining the sought-after outcome. But this approach has caused *at least* vast confusion and unpredictability, if not a range of other harms such as overdeterrence and damaging insurance rates.

My analysis weighs in favor of rejecting the substantial-factor test and returning to a counterfactual approach, even for multiple-sufficient-cause situations. I have argued that a factorial approach to these situations can generally be simplified and applied in practice by using the causal-set approach of the NESS test and the Restatement Third, thus providing an explicit scientific basis for the causal-set approach.

Furthermore, I have shown that my results extend, either directly or by analogy, well beyond tort law and have implications for a wide range of legal contexts in which multiple-sufficient-cause situations arise. These include, for example, criminal law, contracts, and employment discrimination.

Additionally, because the “ordinary” meaning of causal statutory language such as “because of,” “results from,” and “on the basis of” coincides substantially with the scientific meaning of causal language, my results have implications for a wide range of claims involving the judicial interpretation of causal statutory language. These include constitutional claims, criminal claims, employment discrimination claims, and others. In particular, my analysis implies that such terms do not necessarily preclude liability or

unlawfulness in situations in which the alleged cause is accompanied by other sufficient causes.

Finally, as a unique and important application of the foregoing implications, my results imply that, in mixed-motive employment discrimination cases (a specific category of multiple-sufficient-cause situations), the Supreme Court and lower courts should not take for granted that causal language in federal discrimination statutes indicates Congress's intent to exclude from these statutes illegitimate motivating factors (such as race, gender, or age) that are accompanied by sufficient legitimate motivating factors. As I argue in Part V, confusion regarding legislative intent and the meaning of causal language in these statutes has led to vague and potentially inappropriate standards, and a general state of uncertainty in law governing mixed-motive employment discrimination claims. My analysis may help to address this confusion by applying the potential outcomes framework to inform the meaning of such language. This may help to unify a strict construction of causal statutory language on the one hand and intuition regarding the intent of Congress and the purpose of such statutes on the other.

Before concluding, it is necessary to make two notes of caution. First, the factorial approach does not *require* a finding that factors in multiple-sufficient-cause situations constitute "causes." For example, a court may find that, under certain circumstances, Congress intended an interaction-effects analysis rather than a main-effects analysis—i.e., an analysis that conditions explicitly on other factors, such as legitimate motivating factors for an adverse employment action, rather than one that does not. The important point is that an interaction-effects analysis cannot be assumed. Indeed, as I argue in Part IV, in many circumstances, a main-effects analysis involving multiple sets of counterfactuals may be logical and appropriate. Second, it is necessary to distinguish a finding of factual causation from a finding of liability. I do not argue that all factors—even all illegitimate or unlawful factors—in multiple-sufficient-cause situations should necessarily give rise to liability, even when they are appropriately understood as causes under the potential outcomes framework. In many circumstances, courts may appropriately preclude liability based on other legal or policy grounds.

The counterfactual model of causation should not be understood as a narrow test to be applied rigidly to all claims equally regardless of the details of a claim. It is a general framework for asking and answering questions concerning cause and effect. When applied with due nuance to claims involving multiple sufficient causes, the counterfactual model of causation produces results that are consistent with intuition and sound policy, and it permits standards that yield more effective and predictable law.